МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХАРКІВСЬКА НАЦІОНАЛЬНА АКАДЕМІЯ
МІСЬКОГО ГОСПОДАРСТВА

Ачкасов А.Є., Плакіда В.Т., Воронков О.О., Воронкова Т.Б.

ТЕОРІЯ ІМОВІРНОСТЕЙ
І МАТЕМАТИЧНА СТАТИСТИКА

Рекомендовано Міністерством освіти і науки України
як навчальний посібник для студентів вищих навчальних закладів

ХАРКІВ – ХНАМГ – 2008

ББК 22.171я7
УДК 519.2 (075.1)
Т33

Ачкасов А.Є., Плакіда В.Т., Воронков О.О., Воронкова Т.Б. Теорія імовірностей і математична статистика: Навчальний посібник.- Харків, ХНАМГ, 2008.- 247 с.

У навчальному посібнику розглянуті теоретичні та практичні питання з використання елементів теорії імовірностей і математичної статистики. Посібник містить 90 прикладів, велика частка яких має чисельну ілюстрацію розв'язання задач за допомогою вмонтованих функцій пакета MS Excel, що підвищує наочність розв'язання шляхом побудови графіків і гістограм. У посібнику також описані можливості сучасних пакетів прикладних програм і нейронних мереж при розв'язанні задач математичної статистики. Навчальний курс сформовано у системі модульно-рейтингового поетапного контролю знань. У посібнику наведені тестові завдання за змістовими модулями дисципліни, а також варіанти завдань для виконання контрольної роботи.

Навчальний посібник призначений для студентів вищих навчальних закладів, а також може бути корисний для викладачів і фахівців економічних спеціальностей.

Гриф надано Міністерством освіти і науки України,
рішення № 1.4/18-Г-1284 від 04.06.08 р.

Рецензенти:

академік АТН України, віце-президент асоціації економістів-кібернетиків, д.е.н., професор Лисенко Ю.Г. (Донецький національний університет);

д.т.н., професор М.І.Самойленко (Харківська національна академія міського господарства);

д.е.н., професор Г.В.Задорожний (Харківський національний університет ім. В.Н.Каразіна)

ISBN 5-238-00132

ЗМІСТ

ВСТУП

Курс «Теорія імовірностей і математична статистика» є нормативною дисципліною циклу природничо-наукової та загальноекономічної підготовки за освітньо-професійною програмою підготовки бакалавра за спеціальностями напряму 0501 «Економіка і підприємництво». Обсяг курсу становить 108 академічних годин або 3 кредити. Програма курсу розділена на три змістові модулі: «Теорія імовірностей», «Математична статистика» і «Випадкові процеси», відповідно до яких виконується проміжний контроль знань шляхом тестування. Підсумковий контроль знань (залік) проводиться в усній формі. У процесі вивчення курсу студенти повинні виконати контрольну роботу.

Метою вивчення дисципліни «Теорія імовірностей і математична статистика» є формування базових знань в області застосування імовірносно-статистичного апарату, вивчення закономірностей у масових випадкових явищах, визначення їхніх імовірнісних характеристик з метою прогнозування.

У результаті вивчення курсу студенти повинні оволодіти основними методами визначення імовірнісних характеристик випадкових величин, статистичного опису результатів спостереження і перевірки статистичних гіпотез для прийняття на їх основі обґрунтованих рішень.

Теорія імовірностей і математична статистика вивчає закономірності у випадкових явищах. Випадкове явище при багаторазовому повторенні досліду протікає щораз по-іншому. Виникає запитання, чи в кожному випадковому явищі міститься закономірність? Якщо випадкове явище має так звану статистичну однорідність, то воно містить закономірність і називається стохастичним. Стохастичний (греч. stochastikos) — умеющий угадывать. У процесі своєї діяльності людина часто зіштовхується з випадковістю і інтуїтивно припускає наявність у цій випадковості закономірності. Однак іноді покладатися на інтуїцію небажано, все залежить від складності й важливості розв'язуваної проблеми. Тоді виникає необхідність визначити ступінь можливості яких-небудь подій або наслідків шляхом математичних розрахунків. Тут доводиться звертатися до методів теорії імовірностей і математичної статистики.

Зазначимо, що економічні процеси і події є у певній мірі випадковими. Причому, це пов'язане не тільки з «чистою» випадковістю і довільним поводженням людини як елемента економічної системи, а також зі складністю процесів, які піддані впливу такої множини факторів, що не завжди можна врахувати їхній вплив повністю. У зв'язку з цим застосування імовірносно-статистичного апарату в економічних розрахунках є дуже актуальним і ефективним, зокрема при плануванні й прогнозуванні або при оцінці ризику вкладення інвестицій або реалізації бізнес-плану.

Інформація про випадкове явище може бути отримана в результаті його спостереження, тобто шляхом проведення дослідів. Для виявлення закономірності здійснюють обробку дослідних даних. Це завдання вирішує математична статистика. Теоретичною базою останньої є класична теорія імовірностей.

Початок розвитку теорії імовірностей і математичної статистики пов'язаний з європейськими математиками XVII століття. Завдяки роботам швейцарського математика Якоба Бернуллі теорія імовірностей набула найважливіше значення у практичній діяльності людини. Він побудував математичну модель для опису серії незалежних випробувань, довів теорему, яка є окремим випадком закону великих чисел (теорему Бернуллі), що має основне значення в теорії імовірностей і її застосувань до математичної статистики. У XVIII ст. англійський математик Томас Бейес поставив і вирішив одну з основних задач елементарної теорії імовірностей - теорему гіпотез, яка відома під назвою «формула Бейеса». Французький математик П'єр Симон Лаплас розвинув і систематизував результати, отримані Бернуллі. Він довів важливу граничну теорему (теорему Лапласа-Муавра), розвинув теорію помилок, обґрунтував, хоч і нестрого, метод найменших квадратів. Теорія імовірностей у значній мірі сформувалася саме в його роботах. Він ввів теореми додавання і множення імовірностей, поняття математичного сподівання і виробляючих функцій. При його житті робота «Аналітична теорія імовірностей» видавалася тричі. Учень Лапласа, французький математик Сімеон Дені Пуассон ґрунтовно розвинув ідеї Лапласа. Він довів теорему, що торкалася закону великих чисел (закон Пуассона), вперше скориставшись терміном «закон великих чисел».

У XIX ст. теорія імовірностей сформувалася як струнка математична дисципліна у зв'язку з видатними роботами російського математика Пафнутія Львовича Чебишева і його учнів А.М.Ляпунова і А.А.Маркова. П.Л.Чебишев довів загальні форми закону великих чисел. Марков Андрій Андрійович збагатив теорію імовірностей важливими відкриттями й методами. Розвинув метод моментів Чебишева настільки, що стало можливим доведення центральної граничної теореми. Істотно розширив сферу застосування закону великих чисел і центральної граничної теореми, поширивши їх не тільки на незалежні, але й на залежні випробування. Заклав основи однієї із загальних схем природних процесів, які згодом назвали ланцюгами Маркова. Це привело до розвитку нового розділу теорії імовірностей - теорії випадкових процесів. У математичній статистиці А.А.Марков вивів принцип, еквівалентний поняттям незміщених і ефективних статистик. Олександр Михайлович Ляпунов працював у Харківському університеті. Він зробив важливий внесок у теорію імовірностей, давши просте і строге доведення центральної граничної теореми в більш загальній формі, ніж та, в якій вона розглядалася Чебишевим і Марковим. Для доведення він розробив метод характеристичних функцій, який широко застосовується в сучасній теорії імовірностей.

У XX ст. значний внесок у розвиток сучасної теорії імовірностей внесли радянські вчені. А.Н.Колмогоров разом з А.Я.Хінчиним вирішив широко відому систему аксіоматичного обґрунтування, створив теорію марківських процесів з безперервним часом, розвинув теорію стаціонарних випадкових процесів. А.Н.Колмогорову належать дослідження із статистичних методів контролю масової продукції і теорії передачі інформації за каналами зв'язку. Микола Васильович Смирнов отримав фундаментальні результати з розподілу елементів варіаційного ряду та інших питань теорії імовірностей і математичної статистики.

У теорії граничних теорем відом критерій Смирнова. Борис Володимирович Гнеденко розв'язав питання про умови збіжності розподілу сум незалежних доданків до всіх можливих для них розподілів, одержав важливі результати в теорії масового обслуговування і теорії надійності.

Застосування імовірнісних і статистичних методів дає змогу вивчати на науковій основі як діяльність окремих підприємств, так і соціально-економічні процеси в суспільстві в цілому.

Теорія імовірностей і математична статистика є основою для побудови кількісних моделей керування економічними системами. Прикладами таких моделей є моделі планування і керування запасами, теорії ігор, теорії масового обслуговування. Імовірносно-статистичні методи являються базовими для теорії прийняття рішень - складової частини сучасного менеджменту. Статистичні показники аналізують при оцінці ризику в інвестиційній діяльності, в діяльності страхових компаній, а також у багатьох сферах економіки і керування.

Широка сфера застосування теорії імовірностей і математичної статистики обумовлює важливе місце, яке займає даний курс у підготовці економістів вищої кваліфікації, а також менеджерів, маркетологів, бізнесменів, бухгалтерів-аудиторів, соціологів. Ця дисципліна закладає теоретичну основу для наступного вивчення курсів «Математичне програмування», «Економетрія», «Статистика», «Економічний аналіз», «Економічні ризики», «Теорія прийняття рішень».

Цей навчальний посібник підготовлений з метою полегшення сприйняття основних понять теорії імовірностей та методів імовірнісних і статистичних розрахунків. На цій підставі автори не прагнули до повної математичної строгості викладення матеріалу і глибокого теоретичного обґрунтування формул та інших висновків. Для більш глибокого пророблення бажаючі можуть звернутися до спеціальної літератури, вказаної у списку наприкінці посібника. Всі теми проілюстровані достатньою кількістю практичних прикладів, що сприяють засвоєнню відповідних теоретичних положень.

РОЗДІЛ 1
ОПИС ЗМІСТОВИХ РОЗДІЛІВ МОДУЛЯ «ТЕОРІЯ ІМОВІРНОСТЕЙ І МАТЕМАТИЧНА СТАТИСТИКА»

Ціль вивчення модуля «Теорія імовірностей і математична статистика» - формування у студентів базових знань із застосування імовірносно-статистичного апарату.

Завдання вивчення модуля - вивчення математичних і статистичних методів виявлення закономірностей у випадкових явищах.

Предметом модуля «Теорія імовірностей і математична статистика» є кількісні і якісні методи аналізу закономірностей розвитку систем в умовах невизначеності, побудова і аналіз кількісних моделей керування і прогнозування.

Модуль «Теорія імовірностей і математична статистика» належить до циклу дисциплін природничо-наукової та загальноекономічної підготовки відповідно до освітньо-професійної програми підготовки бакалавра за спеціальностями напряму 0501 «Економіка і підприємництво».

Вивчення модуля «Теорія імовірностей і математична статистика» базується на знанні вищої математики і практичних прийомів роботи з табличним процесором Microsoft Excel. Вивчення модуля орієнтоване на освоєння термінології і основних положень теорії імовірностей і математичної статистики, методів обробки статистичного матеріалу, перевірки статистичних гіпотез з метою отримання достовірних результатів.

Кожна тема модуля супроводжується переліком питань для самоконтролю і задачами для самостійного розв'язання.

Виклад матеріалу в посібнику систематизовано відповідно до змістових розділів модуля.

Змістовий модуль 1. ТЕОРІЯ ІМОВІРНОСТЕЙ.

Тема 1. Основні поняття теорії імовірностей.

Основні поняття і визначення теорії імовірностей: випадкові й детерміновані явища, дослід, випадкова подія, імовірність. Класичний і статистичний методи визначення імовірності випадкової події. Поняття про частоту випадкової події.

Тема 2. Операції над подіями. Теореми теорії імовірностей. Основні формули теорії імовірностей.

Сума і добуток випадкових подій. Теорема додавання для несумісних і сумісних подій. Протилежні події. Теорема множення для залежних і незалежних подій. Залежні події. Умовна імовірність випадкової події. Формула повної імовірності. Формула Бейеса (теорема гіпотез). Повторні незалежні випробування. Локальна і інтегральна теореми Лапласа. Формула Пуассона.

Тема 3. Випадкова величина і її закони розподілу

Поняття випадкової величини. Безперервні й дискретні випадкові величини. Закон розподілу випадкової величини. Ряд розподілу. Функція розподілу і її властивості. Щільність розподілу і її властивості.

Тема 4. Числові характеристики випадкової величини

Середнє значення і математичне сподівання. Мода. Медіана. Моменти випадкової величини: початкові й центральні. Дисперсія і середнє квадратичне відхилення. Теореми про числові характеристики.

Тема 5. Найбільш важливі для практики закони розподілу випадкових величин

Біноміальний закон розподілу. Закон розподілу Пуассона. Експонентний закон розподілу. Поняття найпростішого потоку подій. Число подій, що потрапляють на ділянку часу τ. Проміжок часу між двома сусідніми подіями в найпростішому потоці T. Нормальний закон розподілу імовірностей. Інтеграл імовірностей. Правило трьох сигма. Поняття про центральну граничну теорему. Закон рівномірної щільності.

Тема 6. Система випадкових величин. Закони розподілу і числові характеристики системи.

Багатомірна випадкова величина. Поняття системи випадкових величин. Система двох випадкових величин. Функція розподілу і щільність розподілу імовірностей системи двох випадкових величин, їхні властивості. Числові характеристики системи, кореляційний момент і коефіцієнт кореляції. Функції випадкових величин.

Тема 7. Закон великих чисел.

Принцип практичної впевненості. Формулювання закону великих чисел. Рівень значущості. Лема Маркова. Нерівність Чебишева. Теорема Чебишева. Теорема Бернуллі.

Змістовий модуль 2. ЕЛЕМЕНТИ МАТЕМАТИЧНОЇ СТАТИСТИКИ

Тема 8. Обробка статистичних даних.

Вибірковий метод. Поняття генеральної і вибіркової сукупностей. Варіанта. Частота. Варіаційний ряд. Полігон розподілу. Кумулятивна крива. Гістограма розподілу. Визначення закону розподілу спостережуваної ознаки за статистичними даними. Числові характеристики варіаційного ряду. Варіаційний розмах R. Середнє лінійне відхилення d. Дисперсія варіаційного ряду. Стандартне відхилення. Коефіцієнт варіації V. Властивості вибіркових числових характеристик. Довірчий інтервал і довірча імовірність.

Тема 9. Елементи теорії кореляції.

Функціональна, статистична і кореляційна залежності. Рівняння регресії. Поле кореляції. Коефіцієнт регресії. Метод найменших квадратів. Вибірковий коефіцієнт кореляції. Вибіркове кореляційне відношення. Міжгрупова і внутрігрупова дисперсії.

Тема 10. Перевірка статистичних гіпотез.

Поняття статистичної гіпотези. Нульова і альтернативна гіпотези. Критична область. Область прийняття гіпотези. Критична точка zk. Статистичний критерій. Помилки 1-го і 2-го роду. Потужність критерію. Порівняння вибіркової середньої і генеральної середньої нормальної сукупності. t-критерій (розподіл Стьюдента). Порівняння двох дисперсій нормальних генеральних сукупностей. F-критерій (розподіл Фішера). Порівняння вибіркової і генеральної дисперсій нормальної сукупності. Критерій $\chi 2$. Перевірка гіпотези про нормальний розподіл генеральної сукупності. Елементи дисперсійного аналізу. Загальна, факторна і залишкова сума квадратів відхилень. Загальна, факторна і залишкова дисперсії. Елементи регресійного аналізу.

Змістовий модуль 3. ВИПАДКОВІ ПРОЦЕСИ

Тема 11. Елементи теорії випадкових процесів

Поняття випадкового процесу. Імовірнісні характеристики випадкового процесу. Кореляційна функція і її властивості. Лінійне перетворення суми випадкових функцій Xi(t). Стаціонарний випадковий процес. Математичне сподівання і дисперсія стаціонарного випадкового процесу. Кореляційна функція стаціонарного випадкового процесу. Властивість ергодичності. Числові характеристики, визначені за множиною реалізацій і за часом. Елементи теорії масового обслуговування. Основні поняття. Класифікація СМО. Вхідний потік вимог. Дисципліна черги. Механізм обслуговування. Поняття марківського випадкового процесу. Випадковий процес з дискретними станами і безперервним часом. Граф станів системи. Рівняння для імовірностей станів. Одноканальна СМО з відмовами. Фінальні імовірності станів. Характеристики системи масового обслуговування. Багатоканальна СМО з відмовами. Одноканальна СМО з очікуванням і обмеженою чергою. Одноканальна СМО з очікуванням і необмеженою чергою. Багатоканальна СМО з очікуванням і необмеженою чергою. Комп’ютерна обробка статистичних даних.

РОЗДІЛ 2
СКЛАДОВІ МОДУЛЯ «ТЕОРІЯ ІМОВІРНОСТЕЙ І МАТЕМАТИЧНА СТАТИСТИКА»

Змістовий модуль 1. ТЕОРІЯ ІМОВІРНОСТЕЙ.

Тема 1. ОСНОВНІ ПОНЯТТЯ ТЕОРІЇ ІМОВІРНОСТЕЙ

Теорія імовірностей - математична наука, що вивчає закономірності у випадкових явищах. Випадковим називається таке явище, яке при багаторазовому повторенні досліду протікає щораз по-іншому. Наприклад, вимірювання - якщо ми хочемо дістати точний результат, стрілянина по мішені є класичним прикладом випадкового явища, погодні умови, включаючи температуру повітря, силу вітру, атмосферний тиск та ін.

На відміну від випадкових явищ існують детерміновані явища. Це, як правило, закони природи, що вивчаються в курсі фізики, наприклад, прискорення вільного падіння дорівнює 9,8 м/с^2, сила, прикладена до матеріальної точки, додає їй прискорення a F=ma, дохід підприємства пропорційний обсягу продажів Д = Ц * V.

Таким чином, якщо при відтворенні певних умов незмінно відбувається певна подія (та сама, тобто результат незмінно повторюється), то має місце детерміноване явище. Прогноз результату такого досліду можна здійснити, не проводячи експерименту.

У випадку, коли на результат досліду впливає ряд факторів, урахувати які неможливо або дуже складно (крім того, ці фактори можуть змінюватися випадково), скласти математичну модель, що прогнозує розвиток такого явища в детерміністському уявленні неможливо. У такому випадку намагаються знайти у випадкових явищах ті або інші закономірності. Такі закономірності виявляються при масовому повторенні дослідів. Якщо їх вдається знайти, то випадкове явище є статистично однорідним або **стохастичним**. Якщо закономірності в явищі відсутні, тобто виявити їх не вдається, то таке явище є невизначеним і вимагає додаткового дослідження.

Основні поняття і визначення

Одним з основних у теорії імовірностей є поняття випадкової події.

Випадкова подія - це всякий факт, який в результаті досліду може відбутися або не відбутися.

Випадкові події позначають великими літерами латинського алфавіту: A={влучення в мішень}, B={прибуття трамвая на зупинку}, C={поломка технічного пристрою}, D={коротке замикання в мережі}.

Дослідом називається відтворена сукупність умов, у яких може відбутися випадкова подія.

Імовірність випадкової події – це числова міра ступеня об'єктивної можливості появи даної події в результаті досліду. Імовірність події А позначається Р(А).

За одиницю виміру імовірності приймають імовірність **достовірної** події Е, тобто такої, яка в результаті досліду обов'язково відбудеться:

$$P(E) = 1.$$

Протилежна достовірній подія називається **неможливою** і позначається $\overline{E}$. Імовірність неможливої події:

$$P((E) = 0.$$

Очевидно, що імовірність будь-якої випадкової події А міститься між нулем і одиницею:

$$0 \leq P(A) \leq 1.$$

Класичний і статистичний методи визначення імовірності випадкової події

Імовірність випадкової події можна визначити класичним методом тільки в обмеженому числі явищ, а саме, якщо наслідки досліду мають наступні властивості:

утворюють повну групу - якщо результатом однократного випробування є обов'язково один з можливих наслідків;

є рівноможливими - якщо за умови симетрії досліду поява кожного з них є однаково можливою;

є несумісними - якщо будь-які два з них не можуть відбутися одночасно.

Якщо наслідки досліду мають перераховані властивості (утворюють повну групу, є несумісними і рівноможливими), то говорять, що дослід зводиться до **схеми випадків**, або що має місце класична схема теорії імовірностей. У рамках цієї схеми можна точно підрахувати імовірність події, не проводячи випробувань. Якщо дослід зводиться до схеми випадків, то імовірність події А визначається як відношення числа можливих наслідків досліду, які сприяють появі події А, до загального числа можливих наслідків досліду:

$$P(A) = \frac{m}{n}, \tag{1.1}$$

де n - загальне число можливих наслідків досліду; m - число наслідків досліду, які сприяють появі події А.

Щоб підрахувати число всіх випадків n і числа випадків m, які сприяють появі події А, часто використовують число сполучень із s елементів по k елементів:

$$C_s^k = \frac{s!}{k!(s-k)!},$$

де s! = 1*2*3 ... *s, при цьому 0! = 1.

Якщо події в досліді не зводяться до схеми випадків, то оцінку імовірності можна зробити тільки статистично. Спостерігаючи випадкові явища або проводячи випробування, визначають частоту появи даної події. При проведенні серії з n дослідів, у кожному з яких могла з'явитися або не з'явитися подія А, під частотою її появи розуміють відношення

$$P^*(A) = \frac{m}{n} \tag{1.2}$$

де n - число проведених дослідів; m - число появ події А в n дослідах.

Чи можна вважати частоту появи події А його імовірністю? Результат кожного досліду є випадковим, однак якщо спостережуване явище має статистичну однорідність, то при великій кількості дослідів частота події починає стабілізуватися і у границі прагне до імовірності події. Ця властивість усталеності частот, багаторазово перевірена експериментально, і є однією з найбільш характерних закономірностей, спостережуваних у випадкових явищах. Вона відома за назвою закону великих чисел. Бернуллі довів, що

$$\lim_{n\to\infty} P^*(A) = P(A). \tag{1.3}$$

Вираз (1.3) читається так: імовірність події А із збільшенням числа дослідів n зводиться за імовірністю до імовірності події А. Це означає, що із збільшенням числа дослідів n імовірність того, що частота події А відрізняється від імовірності цієї події, зменшується.

Приклад 1.1. В урні знаходяться 3 білих і 4 чорних кулі. Навмання з урни виймають одну кулю. Визначити імовірність того, що вийнята куля опиниться білою.

Розв'язання: Загальне число наслідків досліду дорівнює n=7. Число наслідків досліду, які сприяють появі події A={вийнята куля є білою} дорівнює m=3. Імовірність події А за класичною формулою

$$P(A) = \frac{m}{n} = \frac{3}{7}.$$

Приклад 1.2. З тієї ж урни навмання виймають дві кулі. Визначити імовірність того, що обидві вийнятих кулі виявляться білими.

Розв'язання: Загальне число можливих наслідків досліду можна визначити як число сполучень із 7 елементів по 2 елементи:

$$n = C_7^2 = C_7^2 = \frac{7!}{2!(7-2)!} = 21.$$

Число наслідків досліду, які сприяють появі події B={обидві вийнятих кулі білі}, можна визначити як число сполучень із 3 елементів по 2 елементи:

$$m = C_3^2 = C_3^{2s} = \frac{3!}{2!(3-2)!} = 3.$$

Тоді імовірність події В визначиться за класичною формулою в такий спосіб:

$$P(B) = \frac{m}{n} = \frac{C_3^2}{C_7^2} = \frac{3}{21} = \frac{1}{7}.$$

Приклад 1.3. З тієї ж урни навмання виймають три кулі. Визначити імовірність того, що дві вийнятих кулі чорні, а третя - біла.

Розв'язання: Загальне число можливих наслідків досліду визначимо як число сполучень із 7 елементів по 3 елементи

$$n=C_7^3=\frac{7!}{3!(7-3)!}=35.$$

Число наслідків досліду, які сприяють появі події C={дві вийняті кулі чорні і одна біла}, визначимо в такий спосіб. Число наслідків досліду, які сприяють появі двох чорних куль, дорівнює числу сполучень із 4 елементів по 2 елементи, оскільки дві чорних кулі можуть з'явитися тільки з тих чотирьох чорних, що знаходяться в урні:

$$m_ч=C_4^2=\frac{4!}{2!(4-2)!}=6.$$

Число наслідків досліду, які сприяють появі однієї білої кулі дорівнює числу сполучень із 3 елементів по 1 елементу, оскільки одна біла куля може з'явитися тільки з тих трьох білих, що знаходяться в урні.

$$m_б=C_3^1=C_3^1=\frac{3!}{1!(3-1)!}=3.$$

Тоді імовірність події C

$$P(C)=\frac{m_ч * m_б}{n}=\frac{C_4^2 * C_3^1}{C_7^3}=\frac{6*3}{35}=18/35.$$

Приклад 1.4. Партія виробів містить N виробів, з яких M дефектні. Навмання з партії вибирають k виробів для контролю. Визначити імовірність того, що серед них буде рівно *l* дефектних.

Розв'язання: Зазначимо подію, для якої необхідно визначити імовірність

A={у контрольній партії рівно *l* дефектних виробів}.

Загальне число можливих наслідків досліду дорівнює $n=C_N^k$. Число наслідків досліду, які сприяють появі події A={у контрольній партії рівно *l* дефектних виробів} визначимо в такий спосіб. Число випадків, які сприяють появі *l* дефектних виробів

$$m_д=C_m^l.$$

Число випадків, які сприяють появі k-l не дефектних виробів

$$m_г=C_{N-M}^{k-l}.$$

Імовірність події A

$$P(A)=\frac{m_д * m_г}{n}=\frac{C_m^l * C_{N-M}^{k-l}}{C_N^k}.$$

Запитання для самоперевірки:

1. Дайте визначення випадкової події.
2. Які події називаються: а) достовірними? б) рівноможливими? в) несумісними? г) протилежними? Наведіть приклади.
3. Чи є протилежні події несумісними?
4. Чи є несумісні події протилежними?
5. Дайте визначення імовірності випадкової події.
6. Як підрахувати імовірність події класичним методом?
7. Що розуміють під повною групою подій? Наведіть приклади.
8. Чи завжди можна визначити імовірність випадкової події класичним методом?
9. Як пов'язані між собою імовірність і частота появи події?

Задачі для самостійного розв'язання

1.1. На складі 20 деталей, з яких 17 придатних. Визначити імовірність того, що з трьох навмання взятих деталей всі виявляться придатними.

1.2. На складі 50 придатних і 5 дефектних деталей. Визначити імовірність того, що серед п'яти навмання взятих деталей одна виявиться дефектною.

1.3. Учасники жеребкування тягнуть із ящика жетони з номерами від 1 до 100. Знайти імовірність того, що номер першого жетона, який навмання витягнутий з ящика, не містить цифру 5.

1.4. У партії з 20 готових виробів є 4 бракованих. Партію ділять на дві рівні частини. Визначити імовірність того, що браковані вироби розділяться нарівно.

1.5. Набираючи номер телефону, абонент забув останні дві цифри і, пам'ятаючи, що вони різні, набрав їх навмання. Знайти імовірність того, що були набрані потрібні дві цифри.

1.6. У розіграші першості з баскетболу беруть участь 18 команд, з яких випадковим способом формуються дві групи по 9 команд у кожній. Серед учасників змагань є 5 команд екстракласу. Знайти імовірність того, що а) всі команди екстракласу потраплять в одну групу; б) дві команди потраплять в одну з груп, а три - в іншу.

1.7. Є дві урни, в першій з них a білих і b чорних кулі, у другій – c білих і d чорних. З кожної урни виймають по одній кулі. Знайти імовірність того, що обидві вийнятих кулі виявляться білими.

1.8. У ліфт будинку, в якому сім поверхів, на першому поверсі ввійшли три пасажири. Кожний з них з однаковою імовірністю виходить на кожному з поверхів. Знайти імовірність того, що всі пасажири вийдуть одночасно (на тому самому поверсі).

Тема 2. ОПЕРАЦІЇ НАД ПОДІЯМИ. ТЕОРЕМИ ТЕОРІЇ ІМОВІРНОСТЕЙ. ОСНОВНІ ФОРМУЛИ ТЕОРІЇ ІМОВІРНОСТЕЙ

Оскільки в практичних умовах багаторазове відтворення досліду надзвичайно утруднене, для визначення імовірностей одних випадкових подій по відомих імовірностях інших подій, з ними пов'язаних, користуються теоремами теорії імовірностей: теоремою додавання і теоремою множення.

Введемо визначення.

Сумою двох подій А і В називається подія С, що перебуває в появі події А або події В або обох подій разом:

$$C = A + B.$$

Сума подій - логічна сума, вона називається диз'юнкцією і позначається спеціальним знаком:

$$C = A \cup B.$$

Добутком двох подій А і В називається подія С, що полягає у спільній появі подій А і В:

$$C = A * B.$$

Добуток подій - логічний добуток, він називається кон'юнкцією і також позначається спеціальним знаком:

$$C = A \cap B.$$

Протилежними називають дві несумісних події А і $\overline{A}$, якщо вони складають повну групу.

Подія А називається **незалежною** від події В, якщо імовірність події А не зміниться від того, відбулася подія В чи ні. Якщо ж імовірність події А залежить від того, відбулася подія В чи ні, то такі події називаються **залежними**.

Імовірність події А, обчислена за умови, що подія В мала місце, називається **умовною імовірністю** події А і позначається $P(A|B)$.

Для ілюстрації останнього твердження розглянемо приклад. Нехай в урні три кулі, дві з яких білі, а третя - чорна. Одну за іншою з урни виймають дві кулі. Позначимо події

A={перша вийнята куля виявилася білою};

B={друга вийнята куля виявилася білою}.

Імовірність події В залежить від того, відбулася подія А чи ні. Якщо подія А відбулась, то імовірність події В

$$P(B|A) = \frac{1}{2}.$$

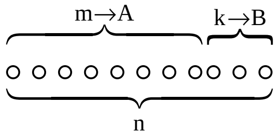
Якщо ж подія А не відбулася, то імовірність події В буде іншою: якщо першою виявилася вийнятою чорна куля, то $P(B|\overline{A}) = 1$.

Теорема додавання

Імовірність суми двох несумісних подій А і В дорівнює сумі імовірностей цих подій, тобто

$$P(A + B) = P(A) + P(B). \tag{2.1}$$

Доведемо це. Нехай дослід має n можливих наслідків, в числі яких m сприяють появі події A і k - появі події B:



Оскільки подія C перебуває в появі події A, якій сприяють m наслідків досліду, або події B, якій сприяють k наслідків, то події C сприяють m+k наслідків досліду. Тоді імовірність події C за класичною формулою визначиться в такий спосіб:

$$P(C) = \frac{m+k}{n} = \frac{m}{n} + \frac{k}{n} = P(A) + P(B).$$

Наслідки теореми додавання:

За методом математичної індукції (узагальнення) теорему додавання імовірностей можна поширити на будь-яке кінцеве число несумісних подій:

$$C = \Sigma A_i$$

$$P(\Sigma A_i) = \Sigma P(A_i).$$

Наслідок 1. Якщо події A_1, A_2, … , A_i, …, A_n утворюють повну групу несумісних подій, то сума їхніх імовірностей дорівнює одиниці:

$$P(\Sigma A_i) = \Sigma P(A_i) = 1. \tag{2.2}$$

Наслідок 2. Сума імовірностей двох протилежних подій дорівнює одиниці:

$$P(A) + P((A) = 1,$$

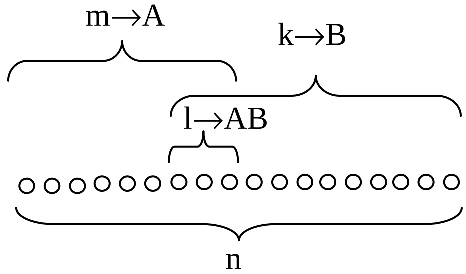
звідки імовірність будь-якої випадкової події

$$P(A) = 1 - P(\bar{A}).$$

Якщо дві події є сумісними, імовірність їхньої суми дорівнює сумі імовірностей цих подій мінус імовірність їхньої спільної появи:

$$P(A + B) = P(A) + P(B) - P(A*B). \tag{2.3}$$

Доведемо це. Нехай дослід має n можливих наслідків, у числі яких m сприяють появі події A, k - появі події B і l - появі події AB:



Оскільки подія C перебуває в появі події A, якій сприяють m наслідків досліду, або події B, якій сприяють k наслідків, то події C сприяють m+k-l наслідків досліду. Тоді імовірність події C за класичною формулою визначиться в такий спосіб:

$$P(C) = \frac{m+k-l}{n} = \frac{m}{n} + \frac{k}{n} - \frac{l}{n} = P(A) + P(B) - P(AB).$$

Можна показати, що для суми трьох сумісних подій справедлива формула

$$P(A+B+C) = P(A)+ P(B)+P(C)- P(AB)- P(AC)- P(BC)- P(ABC),$$

а в загальному випадку для n сумісних подій

$$P\left(\sum_{i=1}^{n} A_i\right) = \sum_i P(A_i) - \sum_{i,j} P(A_i A_j) + \sum_{i,j,k} P(A_i A_j A_k) - \ldots + (-1)^{n-1} P(A_1 A_2 \ldots A_n),$$

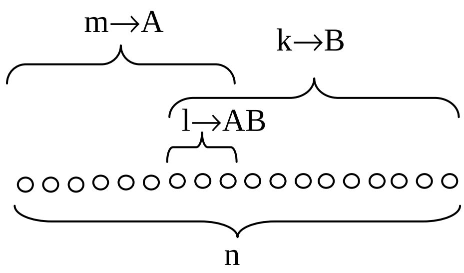
де суми поширюються на всі можливі комбінації індексів i, j, k,..., узятих по одному, по два, по три і т.д.

Теорема множення

Імовірність добутку двох подій A и B дорівнює добутку імовірності одного з них на умовну імовірність іншого, обчислену за умови, що перша відбулася:

$$P(A * B) = P(A) * P(B|A). \qquad (2.4)$$

Доведемо це. Нехай дослід має n можливих наслідків, у числі яких m сприяють появі події A, k - появі події B і l - появі події AB:



Події AB сприяють ***l*** наслідків досліду з n: $P(AB) = \frac{l}{n}$, події A сприяють m наслідків досліду з n: $P(A) = \frac{m}{n}$, а події B, за умови, що одночасно з нею відбулася подія A, сприяють l наслідків досліду з m $P(B|A) = \frac{l}{m}$. Тоді можна записати

$$P(AB) = \frac{l}{n} = \frac{m}{n} * \frac{l}{m}.$$

Остаточно імовірність події добутку двох подій визначиться в такий спосіб:

$$P(AB) = P(A) * P(B/A).$$

Для імовірності добутку n подій формула має вигляд

$$P(A_1 A_2 \ldots A_n) = P(A_1) P(A_2|A_1) P(A_3|A_1 A_2) \ldots P(A_n|A_1 A_2 \ldots A_{n-1}).$$

Якщо події A і B незалежні, то умовна імовірність події B дорівнює безумовній імовірності цієї події:

$$P(B|A) = P(B).$$

Наслідок. Імовірність добутку двох незалежних подій дорівнює добутку імовірностей цих подій:

$$P(A * B) = P(A) * P(B). \qquad (2.5)$$

Якщо маємо кілька незалежних подій:

$$P(\prod_{i=1}^{n} A_i) = \prod_{i=1}^{n} P(A_i).$$

Приклад 2.1. Визначити імовірність пошкодження пристрою, який складається з трьох вузлів. Імовірності пошкодження вузлів відомі і дорівнюють відповідно: $q_1 = 0,02$; $q_2 = 0,01$; $q_3 = 0,001$.

Розв'язання: Позначимо подію, імовірність якої необхідно визначити, A={пристрій пошкоджений}. Виразимо її через елементарні події:

A_1 ={пошкоджений вузол 1};

A_2 ={пошкоджений вузол 2};

A_3 ={пошкоджений вузол 3}.

$$A = A_1 + A_2 + A_3 + A_1 * A_2 + A_2 * A_3 + A_1 * A_3 + A_1 * A_2 * A_3.$$

1-й спосіб. Визначимо імовірність події A за теоремами додавання і множення:

$$P(A) = 0,02+0,01+0,001+0,02*0,01+0,01*0,001+0,02*0,001+$$
$$+0,02*0,01*0,001 = 0,031.$$

2-й спосіб. Визначимо імовірність події A через імовірність протилежної події $\overline{A}$ ={пристрій справний}:

$$\overline{A} = \overline{A}_1 * \overline{A}_2 * \overline{A}_3.$$

Події A_1, A_2, A_3 незалежні, імовірність їхнього добутку дорівнює добутку їхніх імовірностей:

$$P(\overline{A}) = P(\overline{A}_1)*P(\overline{A}_2)*P(\overline{A}_3) = (1-0,02) * (1-0,01) * (1-0,001) = 0,969.$$

Тоді

$$P(A) = 1 - P(\overline{A}_1) = 1-0,969 = 0,031.$$

Приклад 2.2. Студент прийшов здавати іспит, знаючи 15 з 20 питань програми. Визначити імовірність того, що він відповість на три пропонованих екзаменаційних питання.

Розв'язання: Позначимо подію, імовірність якої необхідно визначити, A={студент знає відповіді на три питання}. Виразимо її через елементарні події:

A_1 ={знає перше питання};

A_2 ={знає друге питання};

A_3 ={знає третє питання};

$$A = A_1 * A_2 * A_3.$$

Події A_1, A_2, A_3 – залежні події. Обчислимо їхні умовні імовірності.

$$P(A_1) = \frac{15}{20}.$$

Умовна імовірність події A_2, за умови, що відбулася подія A_1,

$$P(A_2|A_1) = \frac{14}{19}.$$

Умовна імовірність події A_3, за умови, що відбулися події A_1 і A_2,

$$P(A_3|A_1 * A_2) = \frac{13}{18}.$$

Тоді імовірність події А за теоремою множення:

$$P(A) = P(A_1) * P(A_2|A_1) * P(A_3|A_1 * A_2) = \frac{15}{20} * \frac{14}{19} * \frac{13}{18} = 0,368.$$

Приклад 2.3. У шафі знаходяться 9 однакових приладів. На початку досліду вони всі нові. Для тимчасової експлуатації беруть навмання 3 прилади, потім повертають їх у шафу. Таку операцію виконують три рази. Визначити імовірність того, що в шафі залишиться хоча б один новий прилад.

Розв'язання: Позначимо подію, імовірність якої треба визначити A={залишився хоча б один новий прилад}. Меншу кількість варіантів має протилежна подія $\overline{A}$={не залишилося жодного нового приладу}. Така подія може відбутися, тільки якщо вперше і вдруге і втретє для досліду були взяті всі нові прилади. Виразимо її через елементарні події:

B ={в перший раз взяті всі нові прилади};

C ={в другий раз взяті всі нові прилади};

D ={втретє взяті всі нові прилади};

$$A = B * C * D.$$

Імовірність події B ={в перший раз взяті всі нові прилади} дорівнює одиниці:

$$P(B) = 1.$$

Для визначення імовірності події C ={в другий раз взяті всі нові прилади} виразимо її через елементарні події:

C_1 ={перший прилад новий};

C_2 ={другий прилад новий};

C_3 ={третій прилад новий}.

Обчислимо їхні умовні імовірності.

$$P(C_1) = \frac{6}{9}.$$

Умовна імовірність події C_2, за умови, що відбулася подія C_1,

$$P(C_2|C_1) = \frac{5}{8}.$$

Умовна імовірність події C_3, за умови, що відбулися події C_1 і C_2,

$$P(C_3|C_1 * C_2) = \frac{4}{7}.$$

Тоді умовна імовірність події C за теоремою множення:

$$P(C|B) = P(C_1) * P(C_2|C_1) * P(C_3|C_1 * C_2) = \frac{6}{9} * \frac{5}{8} * \frac{4}{7}.$$

Умовну імовірність події D за умови, що відбулися події B і C за теоремою множення визначимо аналогічно:

$$P(D|C*B) = P(D_1) * P(D_2|D_1) * P(D_3|D_1 * D_2) = \frac{3}{9} * \frac{2}{8} * \frac{1}{7}.$$

Тоді імовірність події $\overline{A}$:

$$P(\overline{A}) = P(B) * P(C|B) * P(D|C*B) = 1 * \frac{6}{9} * \frac{5}{8} * \frac{4}{7} * \frac{3}{9} * \frac{2}{8} * \frac{1}{7} = 0{,}003.$$

Імовірність події A={залишився хоча б один новий прилад}:

$$P(A) = 1 - P(\overline{A}_1) = 1\text{-}0{,}003 = 0{,}997.$$

Приклад 2.4. Ви повинні знайти людину, в якої день народження збігається з Вашим. Скількох незнайомців Вам треба буде опитати, щоб імовірність події A={день народження людини збігається з Вашим} була не менше 0,5?

Розв'язання: Імовірність того, що перша людина, в якої Ви запитали, не народилася в один день з Вами,

$$P(\overline{A}) = \frac{365-1}{365}.$$

Якщо Ви опитаєте n людин, то за теоремою множення імовірність, що вони не народилися в один день із Вами:

$$P(\overline{A}) = (\frac{365-1}{365})^n.$$

Тоді імовірність події A={день народження людини збігається з Вашим}:

$$P(A) = 1 - P(\overline{A}) = 1 - (\frac{365-1}{365})^n.$$

Оскільки задана імовірність $P(A) = 0{,}5$, маємо:

$$0{,}5 = 1 - \left(\frac{365-1}{365}\right)^n.$$

Звідки n = 253 чоловік.

Приклад 2.5. У двох урнах знаходяться кулі, що відрізняються тільки кольором. У першій - 5 білих; 11 чорних і 8 червоних, у другій - 10 білих, 8 чорних і 6 червоних. З кожної урни виймають по одній кулі. Визначити імовірність події A={вийняті кулі однакового кольору}.

Розв'язання: Для визначення імовірності події A ={вийняті кулі однакового кольору} виразимо її через елементарні події:

A_1 ={обидві кулі білі};

A_2 ={обидві кулі чорні};

A_3 ={обидві кулі червоні}.

Імовірність події A за теоремою множення для незалежних подій:

$$P(A) = P(A_1) * P(A_2) * P(A_3).$$

Тоді

$$P(A) = 5/24 * 10/24 + 11/24 * 8/24 + 8/24 * 6/24 = 0{,}32.$$

Приклад 2.6. Три спортсмени стріляють по мішені по одному разу. Імовірність влучення для першого з них дорівнює p_1=0,8; для другого p_2=0,85; для третього p_3=0,9. Визначити імовірність подій:

A={всі спортсмени потраплять у мішень};

B={жоден не потрапить};

C={тільки один спортсмен потрапить у мішень};

D={тільки два спортсмени потраплять у мішень};

E={хоча б один спортсмен потрапить у мішень}.

Розв'язання. Визначимо для кожного спортсмена імовірність промаху:

$q_1=1-p_1=1-0,8=0,2;$

$q_2=1-p_2=1-0,85=0,15;$

$q_3=1-p_3=1-0,9=0,1.$

1) Імовірність того, що всі спортсмени потраплять у мішень, знаходимо за теоремою множення імовірностей для незалежних подій.

$$P(A)= p_1*p_2*p_3 = 0,8 * 0,85 * 0,9 =0,612.$$

2) Імовірність того, що жоден із спортсменів не потрапить у мішень, також знаходимо за теоремою множення для незалежних подій

$$P(B)= q_1*q_2*q_3==0,2 * 0,15 * 0,1 =0,003.$$

3) Для обчислення імовірності того, що тільки один спортсмен потрапить у мішень, використаємо теореми додавання для несумісних подій і множення для незалежних подій:

$$P(C)=0,8*0,15*0,1+0,2*0,85*0,1+0,2*0,15*0,9= 0,056.$$

4) Для обчислення імовірності того, що тільки два спортсмени потраплять у мішень, також використаємо теореми додавання для несумісних подій і множення для незалежних подій:

$$P(D)=0,8*0,85*0,1+0,2*0,85*0,9+ 0,8*0,15*0,9= 0,329.$$

Очевидно, що події A, B, C і D складають повну групу. Тоді сума їхніх імовірностей повинна дорівнювати одиниці. Перевіримо це:

$$P(A)+ P(B) + P(C) + P(D)=0,612+0,003+0,056+0,329= 1.$$

5) Подія "хоча б один із спортсменів потрапить у мішень" протилежна події "жоден із спортсменів не потрапить у мішень". Таким чином, маємо

$$P(E) = 1 - P(\bar{A}_1* \bar{A}_2* \bar{A}_3)= 1 - 0,003 =0,997.$$

Формула повної імовірності.

Формула повної імовірності є слідством двох теорем теорії імовірностей. Нехай передбачається проведення досліду, про умови протікання якого можна зробити N взаємовиключних припущень (гіпотез). Умови протікання досліду (гіпотези) являють собою повну групу несумісних подій $H_1, H_2,\ldots, H_N$, імовірності яких $P(H_i)$ відомі. Деяка випадкова подія A може з'явитися при будь-яких умовах протікання досліду з різною імовірністю, тому її можна подати як суму несумісних подій:

$$A=H_1A+H_2A+\ldots+H_NA.$$

Для визначення повної безумовної імовірності події A, використовуючи теореми додавання і множення, дістанемо:

$$P(A) = \sum_{i=1}^{N} P(H_i A) = \sum_{i=1}^{N} P(H_i)P(A / H_i)\,. \qquad (2.6)$$

Таким чином, повна безумовна імовірність події А з урахуванням випадковості умов протікання досліду дорівнює сумі добутків імовірностей кожної з гіпотез на умовну імовірність події А при кожній з гіпотез.

Приклад 2.7. Є три однакові на вигляд урни. У першій - 2 білі і 3 чорних кулі, у другій - 4 білі і 1 чорна, у третій - 3 білі. Навмання з однієї з урн виймають одну кулю. Визначити імовірність того, що вийнята куля виявиться білою.

Розв'язання: Позначимо подію А={вийнята куля біла}. Висуваємо три гіпотези:

H_1 ={обрана перша урна};
H_2 ={обрана друга урна};
H_3 ={обрана третя урна}.

Імовірність кожної гіпотези

$$P(H_1) = P(H_2) = P(H_3) = 1/3.$$

Умовні імовірності події А:

$$P(A|H_1)= \frac{2}{5}; \qquad P(A|H_2)= \frac{4}{5}; \qquad P(A|H_3)= 1.$$

Повна безумовна імовірність події А:

$$P(A)= P(H_1)*P(A|H_1)+P(H_2)*P(A|H_2)+P(H_3)*P(A|H_3) =$$
$$= \frac{1}{3}*\frac{2}{5} + \frac{1}{3}*\frac{4}{5} + \frac{1}{3}*1 = \frac{11}{15}.$$

Приклад 2.8. Пристрій може працювати в одному з трьох режимів: нормальному, перевантаженому і недовантаженому. Нормальний режим спостерігається в 60% випадків; перевантажений - в 30%; недовантажений - в 10%. Імовірність безвідмовної роботи пристрою для нормального режиму дорівнює 0,8, для перевантаженого - 0,5; для недовантаженого - 0,9. Знайти повну з урахуванням випадкових умов імовірність безвідмовної роботи пристрою.

Розв'язання: Позначимо подію А={безвідмовна робота пристрою}. Висуваємо три гіпотези:

H_1 ={пристрій працює в нормальному режимі}
H_2 ={пристрій працює в перевантаженому режимі}
H_3 ={пристрій працює в недовантаженому режимі }

Імовірності гіпотез:

$$P(H_1) = 0,6; \qquad P(H_2) = 0,3; \qquad P(H_3) = 0,1.$$

Умовні імовірності події А:

$$P(A|H_1)= 0,8; \qquad P(A|H_2)= 0,5; \qquad P(A|H_3)= 0,9.$$

Повна безумовна імовірність події А

$$P(A)= P(H_1)*P(A|H_1)+P(H_2)*P(A|H_2)+P(H_3)*P(A|H_3) =$$
$$=0,6*0,8 + 0,3*0,5 + 0,1*0,9 = 0,72.$$

Приклад 2.9. У тирі є п'ять рушниць, імовірність влучення з яких дорівнюють відповідно 0,5, 0,6, 0,7, 0,8, 0,9. Визначити імовірність влучення при одному пострілі, якщо людина бере одну з рушниць навмання.

Розв'язання: Маємо 5 гіпотез, імовірність кожної дорівнює 0,2. Повна безумовна імовірність події А

$$P(A)= P(H_1)*P(A|H_1)+P(H_2)*P(A|H_2)+P(H_3)*P(A|H_3))+$$
$$+P(H_4)*P(A|H_4))+P(H_5)*P(A|H_5) =$$
$$=0{,}2*0{,}5 + 0{,}2*0{,}6 + 0{,}2*0{,}7 + 0{,}2*0{,}8 + 0{,}2*0{,}9 = 0{,}70.$$

Приклад 2.10. Якщо дозволяє погода, пілот саджає літак, користуючись крім приладів, візуальним спостереженням. У цьому разі імовірність нормальної посадки дорівнює 0,99. Якщо над аеродромом низька хмарність, пілот саджає літак, користуючись тільки приладами. У цьому разі імовірність нормальної посадки дорівнює 0,9. Прилади мають надійність 0,8. Якщо над аеродромом низька хмарність і прилади відмовили, то імовірність нормальної посадки дорівнює 0,5. За статистичним даними в 70% випадків над аеродромом низька хмарність. Знайти повну імовірність нормальної посадки.

Розв'язання: Позначимо подію А={нормальна посадка літака}. Гіпотези:

H_1 ={низької хмарності немає};

H_2 ={низька хмарність}.

Імовірності гіпотез H_1 і H_2:

$$P(H_1) = 0{,}3; \qquad P(H_2) = 0{,}7.$$

Умовна імовірність події А при гіпотезі H_1:

$$P(A|H_1) = 0{,}99.$$

Якщо здійсниться гіпотеза H_2, то необхідно висунути ще дві гіпотези:

H_3 ={прилади працюють}

H_4 ={прилади відмовили}

Імовірності цих гіпотез:

$$P(H_3) = 0{,}8; \qquad P(H_4) = 0{,}2.$$

Умовна імовірність події А при гіпотезах H_3 і H_4:

$$P(A|H_3) = 0{,}9; \qquad P(A|H_4) = 0{,}5.$$

Тоді умовна імовірність події А при гіпотезі H_2 може бути визначена за формулою повної імовірності:

$$P(A|H_2) = P(H_3)*P(A|H_3)+P(H_4)*P(A|H_4) = 0{,}8*0{,}9+0{,}2*0{,}5 = 0{,}82.$$

Повна безумовна імовірність події А

$$P(A)= P(H_1)*P(A|H_1)+P(H_2)*P(A|H_2) =$$
$$=0{,}3*0{,}99 + 0{,}7*0{,}82 = 0{,}871.$$

Формула Бейеса (теорема гіпотез)

Ця формула дозволяє за відомими до проведення досліду (апріорним) імовірностями гіпотез $P(H_i)$ і за результатом досліду (настання події А) визначити обчислені після досліду (апостеріорні) імовірності гіпотез $P(H_i|A)$.

За теоремою множення імовірність появи події А при і-й гіпотезі

$$P(H_i*A)= P(H_i)*P(A|H_i)$$

внаслідок симетрії подій справедливо

$$P(H_i*A)= P(A)*P(H_i|A)$$

звідки одержуємо

$$P(H_i / A) = \frac{P(H_i) * P(A / H_i)}{P(A)}$$

або, якщо підставити P(A) з формули (2.6), маємо:

$$P(H_i / A) = \frac{P(H_i) * P(A / H_i)}{\sum_i P(H_i) * P(A / H_i)} \tag{2.7}$$

Таким чином, формула Бейеса дозволяє переоцінити імовірності гіпотез після того, як стає відомим результат досліду, в результаті якого відбулася подія А.

Приклад 2.11. Два стрільці роблять по одному пострілу по мішені. Для першого з них імовірність влучення дорівнює 0,8, а для другого - 0,4. Мішень пробита один раз (одне влучення). Знайти імовірність того, що мішень уражена першим стрільцем.

Розв'язання: Є факт, тобто подія А={мішень уражена один раз}, тобто один із стрільців промахнувся. Висуваємо гіпотези: H_1={мішень уражена першим стрільцем}; H_2={мішень уражена другим стрільцем}. Визначимо імовірності гіпотез. Мішень уражена першим стрільцем, якщо він при пострілі потрапив у мішень, а другий стрілець промахнувся, тоді $P(H_1)$ = 0,8*(1-0,4) = 0,48. Мішень уражена другим стрільцем, якщо він при пострілі потрапив у мішень, а перший стрілець промахнувся, тоді $P(H_2)$ = (1-0,8) *0,4= 0,08. Умовна імовірність події А за умови, що має місце гіпотеза H_1, дорівнює $P(A|H_1)=1$, і за умови, що має місце гіпотеза H_2, дорівнює $P(A|H_1)=1$, тому що в цих випадках мішень буде напевно уражена один раз. Скористаємося теоремою гіпотез і визначимо імовірність реалізації гіпотези H_1:

$$P(H_i / A) = \frac{P(H_i) * P(A / H_i)}{\sum P(H_i) * P(A / H_i)} = \frac{0{,}48*1}{0{,}48*1 + 0{,}08*1} = 0{,}884 .$$

Приклад 2.12. Об'єкт може знаходитися в одному із станів:

H_1 ={працює};

H_2 ={не працює}.

Апріорні імовірності цих станів:

$$P(H_1) = 0{,}7; \qquad P(H_2) = 0{,}3.$$

Є два прилади, що контролюють роботу об'єкта. Вони дають суперечливі повідомлення. Перший прилад повідомляє, що об'єкт працює, а другий – що не працює. На підставі аналізу повідомлень двох приладів визначити імовірності гіпотез H_1 і H_2. Відомо, що перший прилад дає правильні повідомлення з імовірністю p_1 = 0,9, а другий дає правильні повідомлення з імовірністю p_2 = 0,7.

Розв'язання: Є повідомлення двох приладів, тобто подія А={Перший прилад повідомляє, що об'єкт працює, а другий – що не працює}. Імовірності гіпотез задані в умові задачі. Визначимо умовні імовірності події А:

$P(A|H_1)= p_1 * (1 – p_2) = 0,9 * 0,3 = 0,27;$

$P(A|H_2)= (1 - p_1) * p_2 = 0,01 * 0,7 = 0,07.$

Використаємо формулу Бейеса. Апостеріорні імовірності гіпотез:

$$P(H_1 | A) = \frac{P(H_1) * P(A | H_1)}{P(H_1) * P(A | H_1) + P(H_2) * P(A | H_2)} = \frac{0,7 * 0,27}{0,7 * 0,27 + 0,3 * 0,07} = 0,9,$$

$$P(H_2 | A) = \frac{P(H_2) * P(A | H_2)}{P(H_1) * P(A | H_1) + P(H_2) * P(A | H_2)} = \frac{0,3 * 0,07}{0,7 * 0,27 + 0,3 * 0,07} = 0,1,$$

або $P(A|H_2)= 1 - P(A|H_1) =1-0,9 = 0,1$.

Приклад 2.13. З вісімнадцятьох стрільців п'ять потрапляють в мішень з імовірністю 0,8, сім – з імовірністю 0,7, чотири – з імовірністю 0,6 і два – з імовірністю 0,5. Навмання обраний стрілець зробив постріл, але в мішень не потрапив. До якої із груп найімовірніше належав цей стрілець?

Розв'язання: Позначимо факт, що мав місце А={Промах}. Визначимо умовні імовірності цієї події А:

для 1-ї групи стрільців $P(A|H_1)= 0,2;$

для 2-ї групи стрільців $P(A|H_1)= 0,3;$

для 3-ї групи стрільців $P(A|H_1)= 0,4;$

для 4-ї групи стрільців $P(A|H_1)= 0,5.$

$P(A|H_2)= (1 - p_1) * p_2 = 0,01 * 0,7 = 0,07.$

Визначимо імовірності, з якими стрілок, який промахнувся, міг належати до кожної з груп (апріорні імовірності гіпотез):

$P(H_1)= 5/18;$

$P(H_2)= 7/18;$

$P(H_3)= 4/18;$

$P(H_4)= 2/18.$

Визначимо апостеріорні імовірності гіпотез, використовуючи формулу Бейеса:

$$P(H_1 | A) = \frac{P(H_1) * P(A | H_1)}{\sum P(H_i) * P(A | H_i)} = \frac{\frac{5}{18} * 0,2}{\frac{5}{18} * 0,2 + \frac{7}{18} * 0,3 + \frac{4}{18} 0,4 + \frac{2}{18} * 0,5} = \frac{10}{57};$$

$$P(H_2 | A) = \frac{P(H_2) * P(A | H_2)}{\sum P(H_i) * P(A | H_i)} = \frac{\frac{7}{18} * 0,3}{\frac{5}{18} * 0,2 + \frac{7}{18} * 0,3 + \frac{4}{18} 0,4 + \frac{2}{18} * 0,5} = \frac{21}{57};$$

$$P(H_3\,|\,A)=\frac{P(H_3)*P(A\,|\,H_3)}{\sum P(H_i)*P(A\,|\,H_i)}=\frac{\frac{4}{18}*0{,}4}{\frac{5}{18}*0{,}2+\frac{7}{18}*0{,}3+\frac{4}{18}0{,}4+\frac{2}{18}*0{,}5}=\frac{16}{57};$$

$$P(H_4\,|\,A)=\frac{P(H_4)*P(A\,|\,H_4)}{\sum P(H_i)*P(A\,|\,H_i)}=\frac{\frac{5}{18}*0{,}5}{\frac{5}{18}*0{,}2+\frac{7}{18}*0{,}3+\frac{4}{18}0{,}4+\frac{2}{18}*0{,}5}=\frac{10}{57}.$$

Можна виконати перевірку правильності обчислень: $\sum P(H_i|A)=1$. Очевидно, що найбільш імовірною є приналежність стрілка, який промахнувся, до другої групи.

Повторні незалежні випробування

На практиці доводиться зіштовхуватися з такими задачами, які можна уявити у вигляді багаторазово повторюваних незалежних випробувань. Причому імовірність появи події А в одному досліді відома і дорівнює p, а треба визначити імовірність того, що в результаті певного числа дослідів подія А з'явиться рівно m разів. Для визначення цієї імовірності можна використати формулу Бернуллі:

$$P_n(m)=C_n^m p^m q^{n-m}, \tag{2.8}$$

де $P_n(m)$ - імовірність того, що в n випробуваннях подія А з'явиться рівно m разів; C_n^m - число сполучень із n елементів по m; p – імовірність появи події А в одному досліді; q = 1 – p – імовірність не появи події А в одному досліді.

Переконаємося у справедливості формули Бернуллі на простому прикладі. Нехай виконуються п'ять пострілів, імовірність влучення при кожному пострілі в мішень дорівнює p. Визначимо імовірність рівно одного влучення:

$$P_5(1) = pqqqq+qpqqq+qqpqq+qqqpq+qqqqp= 5pq^4;$$

визначимо імовірність рівно двох влучень:

$$P_5(2) = ppqqq+pqpqq+pqqpq+pqqqp+qppqq+$$
$$+qpqpq+qpqqp+qqppq+qqpqp+qqqpp= 10p^2q^3;$$

визначимо імовірність рівно трьох влучень:

$$P_5(3) = pppqq+ppqpq+pqppq+qpppq+ppqqp+$$
$$+pqpqp+qppqp+pqqpp+qpqpp+qqppp= 10p^3q^4;$$

визначимо імовірність рівно чотирьох влучень:

$$P_5(4) = ppppq+pppqp+ppqpp+pqppp+qpppp = 5p^4q^1.$$

У кожному з чотирьох виразів містяться однакові доданки. Число доданків в отриманих виразах для імовірностей визначається як число сполучень з п'яти елементів по одному, по два, по три і по чотири відповідно:

$$C_5^1=\frac{5!}{1!*(5-1)!}=5 \quad C_5^2=\frac{5!}{2!*(5-2)!}=10 \quad C_5^3=\frac{5!}{3!*(5-3)!}=10 \quad C_5^4=\frac{5!}{4!*(5-4)!}=5.$$

Користуючись формулою Бернуллі, можна дістати найімовірніше число появ події А. Найімовірнішим числом появ події А в n незалежних дослідах називається таке число m_0, для якого імовірність перевищує або, принаймні, не менше імовірності кожного з інших можливих чисел появи події А:

$$\begin{matrix} P_n(m_0) \geq P_n(m_0+1) \\ P_n(m_0) \geq P_n(m_0-1) \end{matrix}.$$

Виразимо імовірності в нерівностях за формулою Бернуллі

$$\begin{matrix} \dfrac{n!}{m_0!(n-m_0)!} p^{m_0} q^{n-m_0} \geq \dfrac{n!}{(m_0+1)!(n-m_0-1)!} p^{m_0+1} q^{n-m_0-1} \\ \dfrac{n!}{m_0!(n-m_0)!} p^{m_0} q^{n-m_0} \geq \dfrac{n!}{(m_0-1)!(n-m_0+1)!} p^{m_0-1} q^{n-m_0+1} \end{matrix}.$$

Вирішимо ці нерівності відносно m_0 і одержимо

$$np - q \leq m_0 \leq np + p \qquad (2.9)$$

Причому, m_0 може бути тільки цілим числом. Якщо np - ціле число, то $m_0=np$. Якщо np - q — ціле число, то існують два значення найімовірнішого числа: $m_0 = np - q$ і $m_0' = np - q+1 = np+p$. Якщо np-q — дробове число, то існує одне найімовірніше число: ціле, що міститься між дробовими числами, отриманими з нерівності (2.9).

При дуже великій кількості дослідів n використання формули Бернуллі стає неефективним, тому що обчислення стають громіздкими, а округлення приводять до нагромадження помилки. Асимптотичну формулу дає **локальна теорема Лапласа**, що дозволяє приблизно знайти імовірність появи подій рівно m разів в n дослідах, якщо число випробувань досить велике.

Якщо імовірність p появи події A в кожному випробуванні постійна і відмінна від нуля і одиниці, то імовірність $P_n(m)$ того, що подія A з'явиться в n дослідах рівно m разів, приблизно дорівнює, і тим точніше, чим більше n, значенню функції

$$P_n(m) = \frac{1}{\sqrt{npq}} \varphi(x), \qquad (2.10)$$

де $x = \dfrac{m - np}{\sqrt{npq}}$, а значення функції $\varphi(x) = \dfrac{1}{\sqrt{2\pi}} e^{-x^2/2}$ встановлюються за довідковими таблицями. Функція φ (x) парна, тобто φ(-x) = φ(x).

Якщо в повторних незалежних випробуваннях, в кожному з яких імовірність появи події А постійна і дорівнює p, необхідно обчислити імовірність того, що подія А з'явиться в n випробуваннях не менше m_1 і не більше m_2 разів. Це можна зробити за допомогою **інтегральної теореми Лапласа**.

Якщо імовірність p настання події A в кожному випробуванні постійна і відмінна від нуля і одиниці, то приблизно імовірність $P_n(m_1,m_2)$ того, що подія A з'явиться у випробуваннях від m_1 до m_2 разів,

$$P_n(m_1, m_2) = \frac{1}{\sqrt{2\pi}} \int_{x'}^{x''} e^{-x^2/2} dx \qquad (2.11)$$

де

$$x' = \frac{m_1 - np}{\sqrt{npq}}; \qquad x'' = \frac{m_2 - np}{\sqrt{npq}}$$

При розв'язанні задач користуються спеціальними таблицями, тому що невизначений інтеграл (2.11) не виражається через елементарні функції. Таблиця значень інтегралу $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-x^2/2} dx$ наведена в додатках. Функція Ф(х) непарна, тобто Ф(-х) = -Ф(х). Таблиця містить значення функції Ф(х) тільки для $x \in [0; 5]$; для х > 5 приймають Ф(х) = 0,5.

Тоді приблизно імовірність того, що подія А з'явиться в n незалежних дослідах від m_1 до m_2 разів,

$$P_n(m_1,m_2) = \Phi(x'')-\Phi(x'). \qquad (2.12)$$

Приклад 2.14. У бібліотеці є книги тільки з техніки і математики. Імовірність, що читач візьме книгу з техніки, дорівнює 0,7, а що він візьме книгу з математики - 0,3. Визначити імовірність того, що п'ять читачів підряд візьмуть книги або тільки з техніки, або тільки з математики, якщо кожний з них бере тільки одну книгу.

Розв'язання: Якщо читач бере книгу з техніки з імовірністю 0,7, то п'ять читачів підряд візьмуть книги з техніки з імовірністю

$$P_5(5) = C_5^5 * 0{,}7^5 * 0{,}3^{5-5} = 0{,}168,$$

а п'ять читачів підряд візьмуть книги з математики з імовірністю:

$$P_5(5) = C_5^5 * 0{,}3^5 * 0{,}7^{5-5} = 0{,}00243.$$

Приклад 2.15. Є п'ять однотипних пристроїв. Імовірність безвідмовної роботи кожного дорівнює 0,8. Визначити імовірність того, що в робочому стані перебувають m пристроїв (m = 1, 2, 3, 4, 5).

Розв'язання: Визначимо імовірність, що всі п'ять пристроїв пошкоджені:

$$P_5(0) = C_5^0 * 0{,}8^0 * 0{,}2^{5-0} = 0{,}00032;$$

визначимо імовірність, що чотири пристрої пошкоджені:

$$P_5(1) = C_5^1 * 0{,}8^1 * 0{,}2^{5-1} = 0{,}00638;$$

визначимо імовірність, що три пристрої пошкоджені:

$$P_5(2) = C_5^2 * 0{,}8^2 * 0{,}2^{5-2} = 0{,}0512;$$

визначимо імовірність, що два пристрої пошкоджені:

$$P_5(3) = C_5^3 * 0{,}8^3 * 0{,}2^{5-3} = 0{,}2047;$$

визначимо імовірність, що один пристрій пошкоджений:

$$P_5(4) = C_5^4 * 0{,}8^4 * 0{,}2^{5-4} = 0{,}4095;$$

визначимо імовірність, що жодний пристрій не пошкоджений:

$$P_5(5) = C_5^5 * 0{,}8^5 * 0{,}2^{5-5} = 0{,}328.$$

Σp_i = 0,00032 + 0,00638 + 0,0512 + 0,2047 + 0,4095 + 0,328 = 1.

Приклад 2.16. Імовірність того, що виписаний продавцем чек буде оплачений, дорівнює 0,9. Яке найімовірніше число чеків буде оплачене, якщо виписано 40 чеків?

Розв'язання. Таким чином, n = 40, p=0,9, q=1-0,9=0,1.

1) Скористаємося для розв'язання формулою (2.9) $np - q \le m_0 \le np + p$:

$$40*0{,}9-0{,}1 \le m_0 \le 40*0{,}9 + 0{,}9$$

$$35{,}9 \le m_0 \le 36{,}9.$$

Цій нерівності задовольняє ціле число $m_0 = 36$.

2) Обчислимо добуток np = 40 * 0,9 = 36. Оскільки він є цілим числом, $m_0 = 36$.

Приклад 2.17. Оптова база постачає 10 магазинів, від кожного з яких може надійти заявка на черговий день з імовірністю 0,4, незалежно від заявок інших магазинів. Знайти найімовірніше число заявок в день і імовірність одержання цього числа заявок.

Розв'язання. У цьому випадку n = 10, p=0,4. Добуток np дорівнює цілому числу np = 10*0,4 = 4. Імовірність того, що від магазинів буде отримано 4 заявки, можна визначити за формулою Бернуллі:

$$P_n(m) = C_n^m p^m q^{n-m},$$

де m=4, q=1-0,4=0,6. Тоді імовірність того, що від магазинів буде отримано 4 заявки:

$$P_{10}(4) = C_{10}^4 0{,}4^4 0{,}6^{10-4} = 0{,}251.$$

Приклад 2.18. Визначимо імовірність того, що деяка подія А в 150 дослідах з'явиться рівно 12 разів. Імовірність появи події А в одному досліді дорівнює 0,1.

Розв'язання. У цьому випадку n = 150, p=0,1. Тоді q=1-0,1=0,9, m=12. Для визначення імовірності скористаємося локальною теоремою Лапласа:

$$P_n(m) = \frac{1}{\sqrt{npq}}\varphi(x)$$

обчислимо $x = \frac{m - np}{\sqrt{npq}} = \frac{12 - 150*0{,}1}{\sqrt{150*0{,}1*0{,}9}} = -0{,}82$, значення функції знайдемо за довідковою таблицею φ(-0,82) =0,2939, тоді

$$P_{150}(12) = \frac{1}{\sqrt{150*0{,}1*0{,}9}}*0{,}2939 = 0{,}08.$$

Визначимо імовірність того, що подія А в 150 дослідах з'явиться не менше 10 і не більше 20 разів. Для цього скористуємося інтегральною теоремою Лапласа.

Знайдемо $x' = \frac{m_1 - np}{\sqrt{npq}} = \frac{10 - 150*0{,}1}{\sqrt{150*0{,}1*0{,}9}} = -1{,}36$ і $x'' = \frac{m_2 - np}{\sqrt{npq}} = \frac{14 - 150*0{,}1}{\sqrt{150*0{,}1*0{,}9}} = -0{,}27$.

$$P_n(m_1,m_2) = \Phi(x'')-\Phi(x') = 0{,}4131 - 0{,}1064 = 0{,}307.$$

Формула Пуассона

Якщо імовірність p настання події в окремому випробуванні близька до нуля (p≤0,1), то навіть при великій кількості випробувань n, але при невеликому значенні добутку np (не більше десяти) одержувані за формулою Лапласа значення імовірностей $P_n(m)$ виявляються недостатньо точними і виникає потреба в іншій наближеній формулі. Таким чином, якщо число незалежних випробувань n досить велике, але значення добутку np залишається невеликим, то імовірність того, що в цих випробуваннях подія A наступить m разів, можна визначити за формулою Пуассона:

$$P_n(m) = \frac{(np)^m}{m!} e^{-np} \tag{2.13}$$

Приклад 2.19. Імовірність виготовлення нестандартної деталі дорівнює 0,004. Знайти імовірність того, що серед 1000 деталей буде 5 нестандартних.

Розв'язання: Формалізуємо задачу: n = 1000, p = 0,004, *a* = np = 1000 * 0,004 = 4. Для знаходження імовірності події $P_{1000}(5)$ використаемо формулу Пуассона:

$$P_{1000}(5) = \frac{(4)^5}{5!} e^{-4} = 0{,}1563.$$

Запитання для самоперевірки:

1. Як визначити імовірність суми сумісних подій?
2. Чи може сума двох подій збігатися з їхнім добутком?
3. Наведіть приклади залежних і незалежних подій.
4. Що розуміють під умовною імовірністю події?
5. Як визначається імовірність добутку двох подій?
6. В яких випадках для визначення імовірності застосовується формула Бернуллі?
7. Дайте визначення найімовірнішого числа появ події.
8. Як обчислити найімовірніше число появ події?
9. Чим розрізняються задачі, в яких потрібне застосування локальної і інтегральної граничних теорем?
10. В яких випадках замість формули Бернуллі використовується формула Пуассона?

Задачі для самостійного розв'язання

2.1. На кожній з 5 однакових карток надрукована одна з наступних букв: «О», «П», «Р», «С», «Т». Картки ретельно перемішані. Знайти імовірність того, що на вийнятих по одній і розташованих в одну лінію картках з'явиться слово «СПОРТ».

2.2. В урні *a* білих і b чорних куль. З урни одну за іншою виймають дві кулі. Знайти імовірність того, що обидві кулі будуть білі.

2.3. В урні *a* білих і b чорних куль. З урни одну за іншою виймають всі кулі, що знаходяться в ній. Знайти імовірність того, що другою буде вийнята біла куля.

2.4. Два стрілки незалежно один від іншого роблять два постріли (кожний по своїй мішені). Імовірність влучення в мішень при одному пострілі для першого стрільця p_1, для другого p_2. Виигравшим змагання вважається той стрілець, у мішені якого буде більше пробоїн. Знайти імовірність того, що виграє перший стрілець.

2.5. Імовірність того, що протягом однієї зміни виникне неполадка верстата, дорівнює 0,05. Визначити імовірність того, що протягом трьох змін верстат не зламається жодного разу.

2.6. Прилади одного найменування виготовляють два заводи. На виробництво, що використовує прилади, 2/3 приладів надходять з першого заводу і 1/3 - з другого. Імовірність безвідмовної роботи приладу (надійність), виготовленого першим заводом, дорівнює 0,95, а другим - 0,9. Визначити повну надійність приладу, що надійшов на виробництво.

2.7. Службовці підприємства розподілені за підрозділами і статтю в такий спосіб: у виробничому відділі - 8 жінок і 18 чоловіків, у плановому відділі - 4 жінки і 9 чоловіків, у відділі реалізації - 10 жінок і 5 чоловіків. Навмання обраний службовець виявився чоловіком. Визначити імовірність того, що він працівник а) виробничого відділу; б) планового відділу; в) відділу реалізації.

2.8. Готові вироби містять 5% браку. Визначити імовірність того, що в числі п'яти взятих навмання виробів: а) немає жодного дефектного; б) два дефектних.

2.9. Виробництво дає 1% браку. Визначити імовірність того, що з 1500 виробів бракованих буде не більше 20.

Тема 3. ВИПАДКОВА ВЕЛИЧИНА І ЇЇ ЗАКОНИ РОЗПОДІЛУ

Випадковою називається така величина, яка в результаті досліду може прийняти те або інше значення, невідомо заздалегідь, яке саме (наприклад, число очок, що випали при киданні гральної кістки, число пасажирів у трамваї, температура повітря, час наробітку на відмову технічного пристрою). Випадкову величину позначають прописною літерою латинського алфавіту X, Y або Z, а будь-яке її значення відповідною малою літерою x, y або z.

Розрізняють дискретні й безперервні випадкові величини.

Дискретною називається така випадкова величина, число значень якої скінченне або нескінченне, але рахункове (може приймати тільки окремі значення).

Приклад дискретної випадкової величини - сума очок, що випали при киданні двох гральних кісток, число пасажирів, які проходять через турнікет метро.

Безперервною називається така випадкова величина, число значень якої нескінченне навіть на невеликому інтервалі.

Приклад безперервної випадкової величини - температура повітря. Вона може прийняти кожне з безперервного діапазону значень.

Для повної характеристики випадкової величини треба знати всі можливі її значення, а також імовірності появи цих значень в результаті досліду.

Законом розподілу випадкової величини називається будь-яке правило, яке дозволяє будь-якому значенню випадкової величини поставити у відповідність його імовірність.

Найпростішим законом розподілу є закон розподілу дискретної випадкової величини X, називаний ***рядом розподілу***. Він являє собою таблицю, у верхньому рядку якої перелічені всі значення випадкової величини $x_1, x_2, \ldots, x_n$ в порядку їхнього зростання, а в нижньому - імовірності появи цих значень $p_1, p_2, \ldots, p_n$:

x_i	x_1	x_2	…	x_n
p_i	p_1	p_2	…	p_n

де $p_i = P\{X=x_i\}$.

Оскільки події $\{X=x_1\}$, $\{X=x_2\}$,…,$\{X=x_n\}$несумісні і утворюють повну групу, сума їхніх імовірностей дорівнює одиниці $\sum p_i = 1$ (ця одиниця розподілена між значеннями X). Ряд розподілу може бути заданий і у вигляді графіка (рис. 3.1):

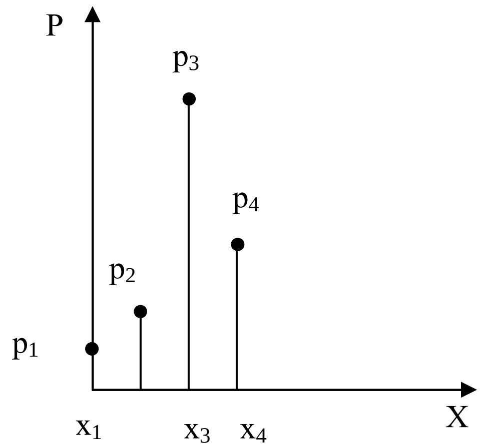


Рис. 3.1 - Графічне зображення ряду розподілу

Найбільш загальною формою закону розподілу для всіх випадкових величин (дискретних і безперервних) є функція розподілу.

Функція розподілу

Для характеристики як безперервних так і дискретних випадкових величин зручніше користуватися не імовірністю події $X=x_i$ (тому що значень x_i може бути багато), а імовірністю того, що випадкова величина X прийняла значення менше свого якогось призначеного x, тобто імовірністю події $X < x$.

Функція розподілу випадкової величини X - це імовірність того, що випадкова величина X прийме значення, менше x:

$$F(x) = P\{X < x\} \tag{3.1}$$

Геометрично функція розподілу - це імовірність того, що значення випадкової величини потрапить лівіше x (рис. 3.2).

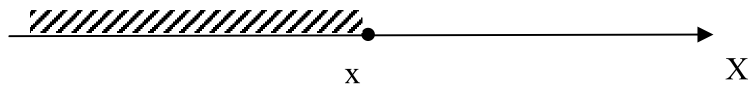


Рис. 3.2 - Розташування значення x на числовій осі

Функція розподілу дискретної випадкової величини являє собою розривну, східчасту функцію, що має перегони в точках, які відповідають можливим значенням $x_1, x_2, \ldots, x_n$ випадкової величини X, які дорівнюють імовірностям $p_1, p_2, \ldots, p_n$ цих значень. Сума всіх стрибків функції розподілу дорівнює одиниці.

У разі безперервної випадкової величини функція розподілу звичайно має вигляд плавної кривої.

Приклад 3.1. Тричі кидають монету. Випадкова величина X - число появ герба. Побудувати ряд розподілу випадкової величини X. Визначити функцію розподілу випадкової величини X и побудувати її графік.

Розв'язання: Побудуємо ряд розподілу X. Очевидно, що число появ герба при триразовому киданні монети може приймати чотири значення 0, 1, 2, 3. Для визначення імовірностей цих значень скористаємося формулою Бернуллі (2.8). Число дослідів n=3, імовірність появи герба в одному досліді p=0,5, імовірність непояви герба в одному досліді q=1-p=1-0,5=0,5. Таким чином, ряд розподілу випадкової величини X має вигляд

x_i	0	1	2	3
p_i	1/8	3/8	3/8	1/8

Виконаємо перевірку: $\sum p_i$ = 1/8+3/8+3/8+1/8=1.

За визначенням функція розподілу випадкової величини X - це імовірність того, що випадкова величина X прийме значення, менше x:

$$F(x) = P\{X < x\}.$$

Обчислимо значення функції розподілу:

$F(0) = P\{X < 0\} = 0;$
$F(1) = P\{X < 1\} = 1/8;$
$F(2) = P\{X < 2\} = 1/8 + 3/8 = 4/8;$
$F(3) = P\{X < 3\} = 4/8 + 3/8 = 7/8;$
при $X \geq 3$ $F(x) = 1.$

Побудуємо графік F(x) (рис. 3.3)

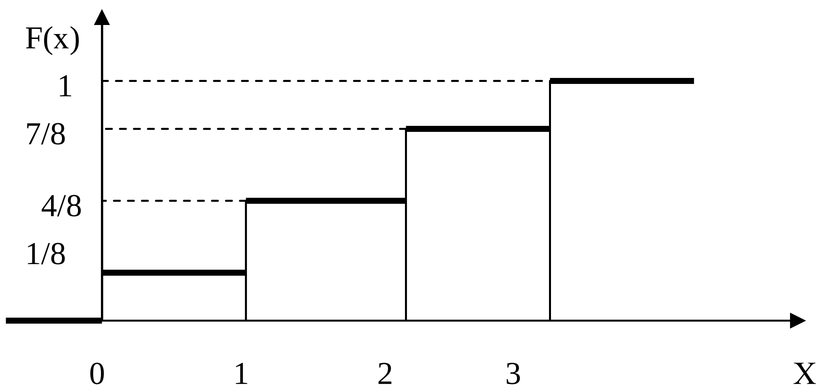


Рис. 3.3 - Графік функції розподілу

Функція розподілу має наступні властивості:

1. Значення функції розподілу належать відрізку [0; 1]

$$0 \leq F(x_2) \leq 1.$$

Це очевидно, оскільки вона є імовірністю.

2. Функція розподілу - неубутна функція, тобто

$$F(x_2) \geq F(x_1), \quad \text{якщо} \quad x_2 > x_1 .$$

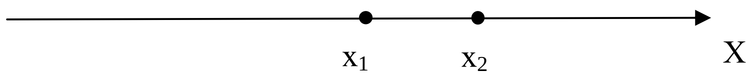


Рис. 3.4 - Розташування значень х на числовій осі

Дійсно, нехай $x_2 > x_1$, тоді функція розподілу $F(x_2)$ буде дорівнювати функції розподілу $F(x_1)$ плюс імовірність влучення X на інтервал значень (x_1, x_2):

$$F(x_2) = F(x_1) + P\{x_1 \le X < x_2\}, \tag{3.2}$$

де $F(x_1) \ge 0$ і $P(x_1 \le X < x_2) \ge 0$, тому що являються імовірностями, отже

$$F(x_2) \ge F(x_1).$$

3. Імовірність того, що випадкова величина X прийме значення, розміщене в інтервалі (x_1, x_2), дорівнює прирісту функції розподілу на цьому інтервалі. З (3.2) маємо:

$$P\{ x_1 \le X \le x_2\} = F(x_2) - F(x_1).$$

4. Якщо всі можливі значення випадкової величини X належать інтервалу $(-\infty, +\infty)$, то при мінус нескінченності функція розподілу дорівнює нулю, а при плюс нескінченності - одиниці, тобто $F(-\infty) = 0$; $F(+\infty) = 1$.

Щільність розподілу

Нехай є безперервна випадкова величина X з функцією розподілу F(x).

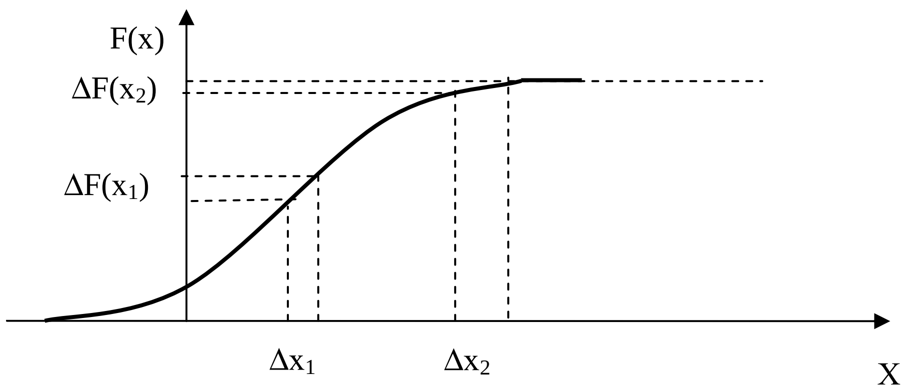


Рис. 3.5 - Функція розподілу безперервної випадкової величини

Говорити про розподіл імовірностей між значеннями безперервної випадкової величини немає рації, тому що число її значень нескінченне навіть на невеликому інтервалі, і імовірність того, що безперервна випадкова величина прийме одне-єдине своє значення x_i, дорівнює нулю. Тому, характеризуючи безперервну випадкову величину, завжди говорять про влучення її значень у той або інший інтервал. З рис. 3.5 видно, що імовірність влучення X в інтервал Δx_1 більше ніж в інтервал Δx_2, оскільки приріст функції розподілу $\Delta F(x_1) > \Delta F(x_2)$. Однак порівнювати прирісти функції розподілу, користуючись її графіком, незручно. З математики відомо, що $\lim\limits_{\Delta x \to 0} \frac{F(x+\Delta x)-F(x)}{\Delta x} = \frac{dF(x)}{dx}$. Таким чином, закон розподілу імовірностей безперервних випадкових величин зручніше визначати завданням не функції розподілу F(x), а щільності розподілу імовірностей f(x), що являє собою похідну від F(x) по x :

$$f(x) = \frac{dF(x)}{dx}. \tag{3.3}$$

Передбачається, що F(x) безперервна і диференційована.

Щільністю розподілу випадкової величини X у точці x називається похідна функції розподілу X в цій точці.

На графіку щільності розподілу імовірність представляється площею (рис. 3.6.).

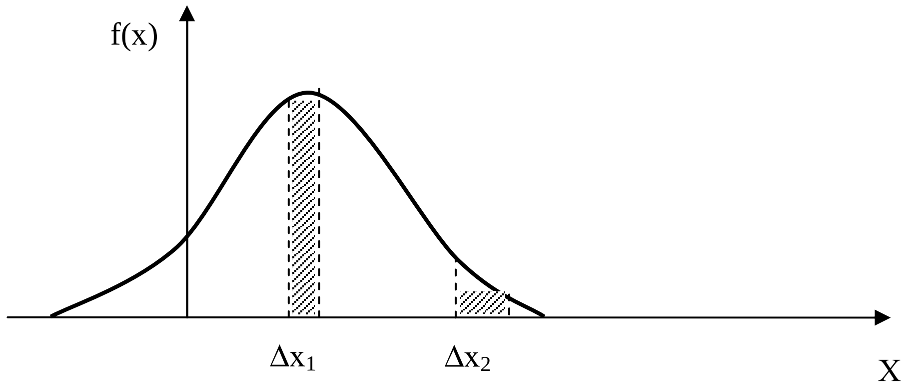


Рис. 3.6 - Графік щільності розподілу

Властивості щільності розподілу імовірностей:

1. Щільність розподілу невід'ємна, тобто f(x) ≥ 0 як похідна неубутної функції.

2. Функція розподілу визначається співвідношенням

$$F(x) = \int_{-\infty}^{x} f(x)dx.$$

3. Інтеграл від щільності розподілу в нескінченних межах дорівнює одиниці:

$$\int_{-\infty}^{+\infty} f(x)dx = 1,$$

де f(x)dx - елемент імовірності, тобто імовірність влучення випадкової величини X на елементарну ділянку dx.

4. Імовірність влучення безперервної випадкової величини в інтервал (x_1, x_2) дорівнює інтегралу щільності розподілу в межах від x_1 до x_2:

$$P\{x_1 \le X \le x_2\} = \int_{x_1}^{x_2} f(x)dx.$$

Функцію розподілу F(x) іноді називають інтегральним законом розподілу, а щільність розподілу f(x) - диференціальним законом розподілу.

Приклад 3.2. Випадкова величина X задана функцією розподілу:

$$F(x)=\begin{cases}0, x\le 0\\ x^2/4, 0<x\le 2\\ 1, x>2\end{cases}$$

Знайти щільність розподілу імовірностей, побудувати графіки функцій f(x) і F(x)

Розв'язання. Знайдемо щільність імовірностей, взявши похідну від функції розподілу на кожному інтервалі:

$$f(x)=\begin{cases}0, x\le 0\\ x/2, 0<x\le 2\\ 0, x>2\end{cases}$$

в) Побудуємо графіки функцій f(x) і F(x).

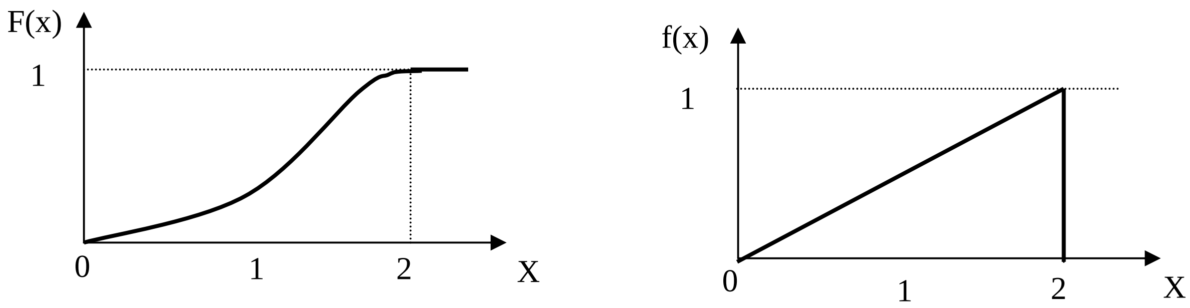


Рис. 3.7 - Графіки функції розподілу і щільності розподілу

Запитання для самоперевірки:

1. Дайте визначення випадкової величини.
2. Яка випадкова величина називається дискретною? Наведіть приклади.
3. Яка випадкова величина називається безперервною? Наведіть приклади.
4. Поясніть, з якою метою в теорії імовірностей розрізняють дискретні і безперервні випадкові величини?
5. Що має на увазі термін «закон розподілу»? В яких формах може бути представлений закон розподілу випадкової величини?
6. Чи може функція розподілу бути: а) більше одиниці; б) від’ємною?
7. Що розуміють під щільністю розподілу випадкової величини?
8. Чому не має сенсу поняття щільності розподілу для дискретної випадкової величини?
9. Яка розмірність щільності розподілу?
10. Перелічіть властивості щільності розподілу.
11. Як, маючи ряд розподілу, знайти значення функції розподілу?
12. Як виражається імовірність влучення випадкової величини на інтервал значень, якщо відомо функцію розподілу? Щільність розподілу?

Задачі для самостійного розв'язання

3.1. Розглядаючи невипадкову величину С як окремий вид випадкової, побудувати для неї функцію розподілу, знайти математичне сподівання і дисперсію.

3.2. У партії з 30 виробів є 7 дефектних. З цієї партії випадковим способом обрані три вироби для перевірки їхньої якості. Побудувати ряд розподілу числа відібраних для перевірки виробів (випадкової величини X).

3.3. Для умов попередньої задачі побудувати функцію розподілу числа відібраних для перевірки виробів (випадкової величини X).

3.4. Випадкова величина X задана функцією розподілу

$$F(x)=\begin{cases}0, & x<2\\ (x-2)^2, & 2\le x\le 3\\ 1, & x>3\end{cases}$$

Знайти щільність розподілу f(x) і визначити імовірність влучення випадкової величини X в інтервал (1,5; 2).

Тема 4. ЧИСЛОВІ ХАРАКТЕРИСТИКИ ВИПАДКОВОЇ ВЕЛИЧИНИ

Закон розподілу випадкової величини являє собою деяку функцію, що повністю описує випадкову величину з імовірнісної точки зору, тобто є її вичерпною характеристикою і дозволяє визначати імовірності будь-яких подій, пов'язаних з випадковою величиною. Однак у багатьох практичних задачах потрібно отримати більш компактне уявлення про випадкову величину. Для теорії імовірностей і її додатків велику роль відіграють деякі постійні числа, одержувані за певними правилами із законів розподілу випадкових величин і називані **числовими характеристиками** випадкової величини. Найважливішою числовою характеристикою випадкової величини є **математичне сподівання.**

Математичним сподіванням випадкової величини X називається сума добутків всіх можливих її значень на імовірності цих значень

$$M[X]=\sum_{i=1}^{n} x_i p_i \,. \tag{4.1}$$

Математичне сподівання тісно пов'язане із середнім значенням випадкової величини, отриманим з великої кількості дослідів.

Нехай зроблено n незалежних дослідів, в кожному з яких X прийняла певні значення x_i, $i=\overline{1,k}$. Припустимо, що x_1 з'явилося n_1 разів, x_2 – n_2 разів, і т.д. і $\sum_{i=1}^{k} n_i = n$. Знайдемо середнє арифметичне отриманих значень, позначивши його m*:

$$m^*=\frac{x_1n_1+x_2n_2+\ldots+x_kn_k}{n}.$$

Очевидно, що $\frac{n_i}{n}$ - частота (статистична імовірність) події {X=xi} p*, тоді:

$$m^* = \sum_{i=1}^{k} x_i p^*_i \ ,$$

тобто середнє арифметичне дорівнює сумі добутків значень випадкової величини на їхні частоти.

При збільшенні числа дослідів n частота події p_i* збігається за імовірністю до імовірності події p_i. Виходить, і середнє арифметичне m* буде збігатися за імовірністю до математичного сподівання M[X].

Таким чином, математичне сподівання характеризує середнє значення випадкової величини. Воно є характеристикою її положення на числовій осі, навколо якого групуються всі можливі значення. Іншими характеристиками положення є мода і медіана.

Модою Мо дискретної випадкової величини називається найбільш імовірне її значення. Для безперервної випадкової величини модою Мо є значення, якому відповідає найбільше значення її щільності розподілу, іншими словами, мода - точка глобального максимуму кривої розподілу безперервної випадкової величини (рис. 4.1).

Медіаною M_e випадкової величини називається таке її значення, що поділяє пополам площу, обмежену кривою розподілу, тобто

P{X<Me}=P{X>Me}=0,5.

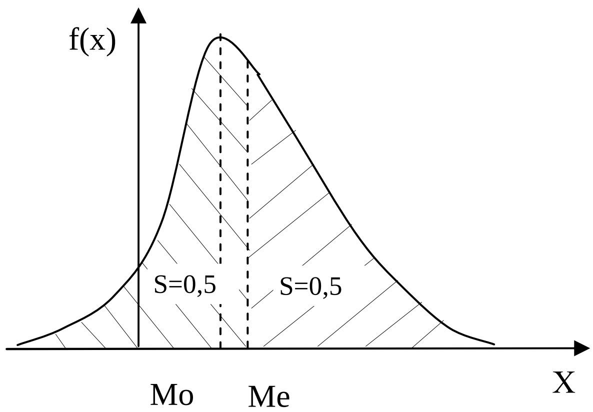


Рис. 4.1 - Мода Мо і медіана Me

Узагальненням числових характеристик випадкової величини є так звані **моменти** або математичні сподівання випадкової величини. Розрізняють початкові α і центральні μ моменти.

Початковим моментом s -го порядку дискретної випадкової величини називається математичне сподівання s-ї степені цієї величини:

$$\alpha_s = \sum_{i=1}^{n} x_i^s p_i \ . \tag{4.2}$$

Для безперервної випадкової величини початковий момент s -го порядку

$$\alpha_s = \int_{-\infty}^{+\infty} x^s f(x)dx \ . \tag{4.3}$$

Вирази (4.2) і (4.3) можна об'єднати в один, користуючись знаком математичного сподівання *M:*

$$\alpha_s = \mathrm{M}[\,X^s\,]\,, \tag{4.4}$$

тобто початковим моментом s -го порядку випадкової величини називається математичне сподівання s -ї степені цієї випадкової величини.

При s = 1 дістаємо перший початковий момент або математичне сподівання випадкової величини:

$$\alpha_1 = \mathrm{M}[\,X\,] = m_x\,. \tag{4.5}$$

На практиці іноді застосовують другий початковий момент α_2:

$$\alpha_2 = \mathrm{M}[\,X^2\,]\,. \tag{4.6}$$

Центральним моментом s-го порядку випадкової величини X називається математичне сподівання s -ї степені центрованої величини X. Під центрованою розуміється відхилення випадкової величини від її математичного сподівання:

$$\overset{0}{X} = X - m_x\,.$$

Центрування випадкової величини рівнозначне переносу початку відліку її значень.

Моменти центрованої випадкової величини називаються центральними. Центральний момент s-го порядку випадкової величини X виражається формулою

$$\mu_s = \mathrm{M}[\,\overset{0}{X}{}^{s}\,]\,. \tag{4.7}$$

Для дискретної випадкової величини X:

$$\mu_s = \mathrm{M}[\,\overset{0}{X}{}^{s}\,] = \sum_{i=1}^{n} \overset{0}{x}{}_i^{s}\, p_i = \sum_{i=1}^{n} (\,x_i - m_x\,)^s\, p_i\,. \tag{4.8}$$

Для безперервної випадкової величини X:

$$\mu_s = \mathrm{M}[\,\overset{0}{X}{}^{s}\,] = \int_{-\infty}^{+\infty} \overset{0}{x}{}^{s}\, f(\,x\,)dx_i = \int_{-\infty}^{+\infty} (\,x - m_x\,)^s\, f(\,x\,)dx\,. \tag{4.9}$$

Центральний момент першого порядку дорівнює нулю:

$$\mathrm{M}[\overset{0}{\mathrm{X}}] = \mathrm{M}[\mathrm{X} - \mathrm{m_x}] = \sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{x_i} - \mathrm{m_x})\mathrm{p_i} = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{x_i p_i} - \mathrm{m_x} \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathrm{p_i} = 0\,.$$

Центральний момент другого порядку для дискретної випадкової величини X:

$$\mu_2 = \mathrm{M}[\overset{0}{\mathrm{X}}{}^{2}] = \sum_{\mathrm{i}=1}^{\mathrm{n}} \overset{0}{\mathrm{x}}{}_{\mathrm{i}}^{2}\, \mathrm{p_i} = \sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{x_i} - \mathrm{m_x})^2 \mathrm{p_i} = \mathrm{D_x}\,. \tag{4.10}$$

Для безперервної випадкової величини:

$$\mu_2 = \mathrm{M}[\overset{0}{\mathrm{X}}{}^{2}] = \int_{-\infty}^{+\infty} \overset{0}{\mathrm{x}}{}^{2}\, \mathrm{f(x)dx} = \int_{-\infty}^{+\infty} (\mathrm{x} - \mathrm{m_x})^2 \mathrm{f(x)dx} = \mathrm{D_x}\,. \tag{4.11}$$

Другий центральний момент називається **дисперсією**. Дисперсія випадкової величини є характеристикою розсіювання цієї величини навколо математичного сподівання. У випадку якщо це розсіювання відсутнє, величина D_x дорівнює нулю. Дисперсія має розмірність квадрата випадкової величини, що не завжди зручно. Тому як характеристику розсіювання часто використовують середнє квадратичне відхилення X:

$$\sigma_x = \sqrt{D_x}. \tag{4.12}$$

Центральні моменти більш високого порядку можуть характеризувати ступінь асиметрії розподілу випадкової величини, крутість кривої розподілу та ін.

Приклад 4.1. Визначимо числові характеристики дискретної випадкової величини для умов прикладу 3.1. Маємо ряд розподілу:

x_i	0	1	2	3
p_i	1/8	3/8	3/8	1/8

Розв'язання: Визначимо математичне сподівання випадкової величини X:

$$m_x = \Sigma x_i{*}p_i = 0{*}1/8 + 1{*}3/8 + 2{*}3/8 + 3{*}1/8 = 1{,}5.$$

Дисперсію визначимо двома способами:

за формулою другого центрального моменту:

$$D_x = \Sigma(x_i-m_x)^2{*}p_i = (0-1{,}5)^2{*}1/8 + (1-1{,}5)^2{*}3/8 + (2-1{,}5)^2{*}3/8 + (3-1{,}5)^2{*}1/8 = 0{,}75.$$

за формулою, що містить другий початковий момент α_2:

$$D_x = \alpha_2 - m_x^2$$

$$\alpha_2 = \Sigma x_i^2{*}p_i = 0^2{*}1/8 + 1^2{*}3/8 + 2^2{*}3/8 + 3^2{*}1/8 = 3.$$

$$D_x = \alpha_2 - m_x^2 = 3-1{,}5^2 = 0{,}75.$$

Визначимо середнє квадратичне відхилення випадкової величини X:

$$\sigma_x = \sqrt{D_x} = \sqrt{0{,}75} = 0{,}855.$$

Приклад 4.2. Нехай задано ряд розподілу.

x_i	1,4	1,8	2,3	3,2
p_i	0,3	0,4	0,2	0,1

Визначити числові характеристики випадкової величини X.

Розв'язання: Математичне сподівання випадкової величини X:

$$M[X]=x_1{*}p_1+x_2{*}p_2+x_3{*}p_3+x_4{*}p_4=$$
$$= 1{,}4 * 0{,}3 + 1{,}8{*}0{,}4 + 2{,}3 * 0{,}2 + 3{,}2 * 0{,}1 = 1{,}92.$$

Дисперсія:

$$D[X] = (x_1 - M[X])^2 * p_1 + (x_2 - M[X])^2 * p_2 + (x_3 - M[X])^2\, p_3 +$$
$$+ (x_4{*} - M[X])^2 p_4 = (1{,}4-1{,}92)^2 * 0{,}3 + (1{,}8 -1{,}92)^2 * 0{,}4 + (2{,}3-1{,}92)^2 * 0{,}2 +$$
$$+(3{,}2-1{,}92)^2 * 0{,}1 = 0{,}28.$$

або $$M[X^2]= 1{,}4^2 * 0{,}3 + 1{,}8^2 * 0{,}4 + 2{,}3^2 * 0{,}2 + 3{,}2^2 * 0{,}1 = 3{,}966;$$

$$D[X] = M[X^2] – (M[X])^2 = 3{,}966 – 1{,}92^2 = 0{,}28.$$

Середнє квадратичне відхилення випадкової величини X:

$$\sigma[X] = \sqrt{D[X]} = \sqrt{0,28} = 0,53.$$

Приклад 4.3. Нехай випадкова величина X задана функцією розподілу:

$$F(x) = \begin{cases} 0, x \le 0 \\ x^2/4, 0 < x \le 2 \\ 1, x > 2 \end{cases}.$$

Визначимо числові характеристики безперервної випадкової величини.

Розв'язання: Знайдемо щільність імовірностей:

$$f(x) = \begin{cases} 0, x \le 0 \\ x/2, 0 < x \le 2 \\ 0, x > 2 \end{cases}.$$

Визначимо математичне сподівання X:

$$M[X] = \int_0^2 x * x/2dx = 1/2\int_0^2 x^2dx = \frac{1}{2}\frac{x^3}{3}\Big|_0^2 = 4/3.$$

Знайдемо дисперсію X:

$$D[X] = \int_0^2 x^2 * x/2dx - (4/3)^2 = 1/2\int_0^2 x^3dx - (4/3)^2 = \frac{1}{2}\frac{x^4}{4}\Big|_0^2 - (4/3)^2 = 2/9.$$

Властивості числових характеристик

1. Математичне сподівання невипадкової величини *c* дорівнює їй самій:

$$M[c] = \sum_{i=1}^{n} cp_i = c\sum_{i=1}^{n} p_i = c. \qquad (4.13)$$

2. Математичне сподівання добутку невипадкової величини *c* на випадкову величину X дорівнює добутку цієї невипадкової величини на математичне сподівання випадкової величини X:

$$M[cX] = \sum_{i=1}^{n} cx_i p_i = c\sum_{i=1}^{n} x_i p_i = cM[X], \qquad (4.14)$$

тобто невипадкову величину c можна виносити за знак математичного сподівання.

3. Дисперсія невипадкової величини *c* дорівнює нулю:

$$D[c] = M[\overset{0}{c}{}^2] = M[(c - m_c)^2] = M[0] = 0. \qquad (4.15)$$

4. Дисперсія добутку невипадкової величини *c* на випадкову величину X дорівнює добутку квадрата цієї невипадкової величини на дисперсію випадкової величини X:

$$D[cX] = M[c^2 \overset{0}{X}{}^2] = c^2M[\overset{0}{X}{}^2] = c^2D[X] \qquad (4.16)$$

тобто невипадкову величину можна виносити з-під знака дисперсії, звівши її у квадрат.

На практиці дисперсію часто обчислюють як різницю другого початкового моменту α_2 і квадрата математичного сподівання:

$$D[X]=M[\overset{0}{X}{}^2]=M\left[(X-M[X])^2\right]=M\left[X^2-2XM[X]+(M[X])^2\right]$$

оскільки математичне сподівання суми незалежних випадкових величин дорівнює сумі їхніх математичних сподівань, а постійний множник можна виносити за знак математичного сподівання, маємо:

$$D[X]=M[X^2]-2M[X]M[X]+(M[X])^2=M[X^2]-(M[X])^2 \qquad (4.17)$$

Запитання для самоперевірки:

1. Назвіть основні числові характеристики випадкових величин.
2. Як пов'язані між собою математичне сподівання і середнє арифметичне значень випадкової величини?
3. Математичне сподівання - випадкова величина чи ні?
4. Чи є дисперсія випадковою величиною?
5. Як математичне сподівання і дисперсія характеризують випадкову величину?
6. Чим зручне застосування замість дисперсії середнього квадратичного відхилення?
7. В яких одиницях вимірюють математичне сподівання?
8. В яких одиницях вимірюють дисперсію?
9. Чому дорівнює математичне сподівання невипадкової величини С?
10. Як мода і медіана характеризують випадкову величину?

Задачі для самостійного розв'язання

4.1. Для умов прикладу 3.2 визначити математичне сподівання, дисперсію і середнє квадратичне відхилення випадкової величини.

4.2. Для умов прикладу 3.4 визначити математичне сподівання, дисперсію і середнє квадратичне відхилення випадкової величини.

4.3. До випадкової величини X додали невипадкову величину *a*. Як від цього зміняться її числові характеристики: математичне сподівання, дисперсія і середнє квадратичне відхилення?

4.4. Випадкову величину X помножили на невипадкову величину *a*. Як від цього зміняться її числові характеристики: математичне сподівання, дисперсія і середнє квадратичне відхилення?

4.5. Виконують два незалежних постріли по мішені. Імовірність влучення при кожному пострілі дорівнює p. Розглядаються дві випадкові величини: X-різниця між числом влучень і числом промахів і Y - сума числа влучень і числа промахів. Побудувати для випадкових величин X і Y ряд розподілу (для кожної окремо) і знайти їхні числові характеристики.

Тема 5. НАЙБІЛЬШ ВАЖЛИВІ ДЛЯ ПРАКТИКИ ЗАКОНИ РОЗПОДІЛУ ВИПАДКОВИХ ВЕЛИЧИН

Біноміальний закон розподілу

Дискретна випадкова величина X має біноміальний закон розподілу (розподіл Бернуллі), якщо її можливі значення: 0, 1, ..., n, а відповідні імовірності визначаються співвідношенням

$$P_n(m) = C_n^m p^m q^{n-m} \tag{5.1}$$

де p - імовірність появи події A в одному досліді, 0<p<1; q - імовірність не появи події A в одному досліді, q=1-p.

Таким чином, розподіл залежить від двох параметрів n і p.

Для визначення числових характеристик випадкової величини, що має біноміальний закон розподілу, введемо поняття **виробляючої функції.**

Нехай дискретна випадкова величина X приймає невід'ємні цілочисельні значення 0, 1, 2,…, k з імовірностями p_1, p_2, …, p_k (p_k=P{X=k})...
Виробляючою функцією для випадкової величини X називається вираз вигляду:

$$\varphi(z) = \sum_{k=0}^{\infty} p_k z^k \tag{5.2}$$

де z – довільний параметр, $0<z\le 1$. Коефіцієнти p_k при z^k у виробляючій функції дорівнюють імовірностям того, що випадкова величина X прийме значення k. Вираз (5.2) залишається справедливим і якщо число значень X скінчене, тому що при k>n імовірності p_k обертаються в нуль.

При z=1 виробляюча функція дорівнює одиниці:

$$\varphi(1) = \sum_{k=0}^{\infty} p_k = 1. \tag{5.3}$$

Візьмемо похідну по z від виробляючої функції:

$$\varphi'(z) = \sum_{k=0}^{\infty} k p_k z^{k-1}.$$

Нехай z=1, тоді $\varphi'(1) = \sum_{k=0}^{\infty} k p_k$, тобто похідна по z від виробляючої функції при z=1 являє собою суму добутків значень X на їхній імовірності, а отже є математичним сподіванням X.

$$M[X]=m_x=\varphi'(1). \tag{5.4}$$

Візьмемо другу похідну від φ(z):

$$\varphi''(z) = \sum_{k=0}^{\infty} k(k-1) p_k z^{k-2}.$$

Нехай z=1, дістанемо $\varphi''(1) = \sum_{k=0}^{\infty} (k^2 - k) p_k = \sum_{k=0}^{\infty} k^2 p_k - \sum_{k=0}^{\infty} k p_k$. Перший доданок – другий початковий момент α_2, а другий - математичне сподівання X. Одержимо вираз другого початкового моменту α_2.:

$$\alpha_2=\varphi''(1) + \varphi'(1). \qquad (5.5)$$

Тобто другий початковий момент дорівнює сумі першої і другої похідних виробляючої функції при z=1.

Визначимо числові характеристики випадкової величини X, розподіленої за біноміальним законом. Запишемо виробляючу функцію:

$$\varphi(z) = \sum_{m=0}^{n} P_m z^m = \sum_{m=0}^{n} C_n^m p^m q^{n-m} z^m .$$

Такий же вигляд має n-я степінь бінома:

$$\varphi(z)=(q+pz)^n. \qquad (5.6)$$

Візьмемо похідну від виразу (5.6):

$$\varphi'(z)= np(q+pz)^{n-1}$$

і підставимо z=1, дістанемо математичне сподівання X:

$$m_x=\varphi'(1) = np(q+p)^{n-1}= np(1)^{n-1}= np. \qquad (5.7)$$

Для обчислення дисперсії знайдемо другий початковий момент:

$$\alpha_2=\varphi''(1) +m_x$$

$$\varphi''(z)= n(n-1)p^2(q+pz)^{n-2}; \qquad \varphi''(1) = n(n-1)p^2$$

$$\alpha_2= n(n-1)p^2+ np.$$

$$D_x=\alpha_2-m_x^2= n(n-1)p^2+ np - n^2p^2 = n^2p^2 - np^2 + np - n^2p^2 = np-np^2 = np(1-p)=$$
$$=npq.$$

Таким чином, числові характеристики випадкової величини, розподіленої за біноміальним законом, мають вигляд:

$$m_x = np; \qquad D_x = npq; \qquad \sigma_x = \sqrt{npq}. \qquad (5.8)$$

Приклад 5.1. У відділ верхнього одягу універмагу один за іншим входять три відвідувачі. За оцінками менеджера імовірність того, що відвідувач, який ввійшов, зробить покупку, дорівнює 0,3. Визначити імовірність того, що: а) жоден з відвідувачів нічого не купить; б) тільки один відвідувач зробить покупку; в) два відвідувачі зроблять покупку; г) всі троє куплять що-небудь у відділі. Побудувати ряд розподілу і визначити математичне сподівання і дисперсію випадкової величини X - число відвідувачів, які зробили покупку.

Розв'язання. Поява трьох відвідувачів у відділі універмагу може розглядатися як проведення трьох дослідів. Досліди однакові, тому що імовірність появи події A={здійснення покупки одним відвідувачем} однакова для всіх трьох і дорівнює p=0,3. Відповідно імовірність непокупки для кожного відвідувача q=0,7. Наслідки дослідів незалежні, тому що рішення про покупку для кожного з відвідувачів не залежить від рішень інших відвідувачів відділу.

Для визначення імовірностей біноміального розподілу випадкової величини X скористаємося формулою (5.1).

x_i	0	1	2	3
p_i	0,343	0,441	0,189	0,027

Обчислимо математичне сподівання і дисперсію за формулами моментів випадкової величини:

$$m_x = 0*0,343+1*0,441+2*0,189+3*0,027 = 0,9$$

$$D_x = 0*(0,343-0,9)^2+1*(0,441-0,9)^2+2*(0,189-0,9)^2+3*(0,027-0,9)^2 = $$
$$=0,63.$$

Обчислимо математичне сподівання і дисперсію за формулами (5.8):

$$m_x = n*p = 3 * 0,3 = 0,9$$

$$D_x = n*p*q = 3 * 0,3 * 0,7 = 0,63.$$

Побудуємо графік отриманого розподілу (рис. 5.1).

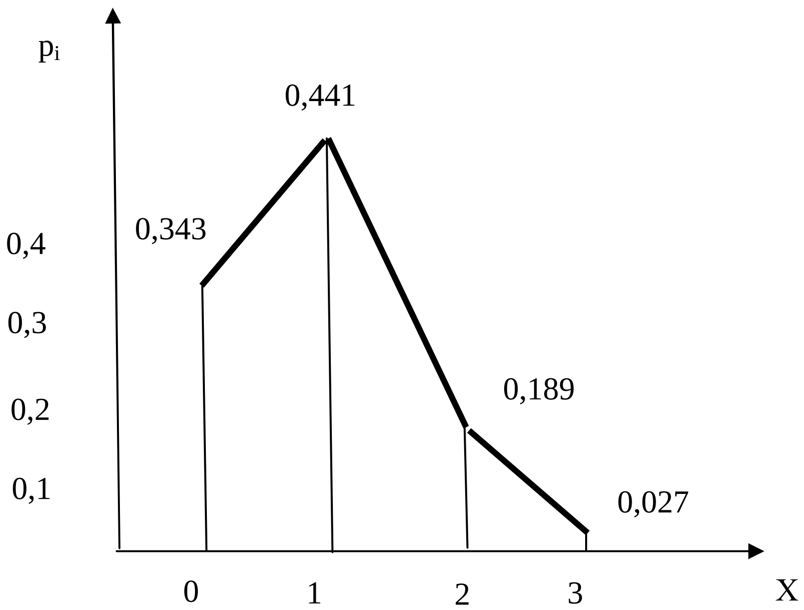


Рис. 5.1 - Многокутник розподілу

Закон розподілу Пуассона

Закон Пуассона є граничним для біноміального розподілу. Тобто він має місце, коли n нескінченно велике, а p дуже мала (його називають законом рідких явищ).

Імовірність того, що за якийсь час τ відбудеться рівно k подій за законом Пуассона визначається формулою

$$P(k) = \frac{(\lambda\tau)^k}{k!} e^{-\lambda\tau}, \tag{5.9}$$

де λ - число подій в одиницю часу; τ - інтервал часу.

Покажемо, що коли $n \to \infty$, а p дуже мала, біноміальний розподіл можна приблизно заміняти пуассонівським.

Розділимо інтервал τ на n ділянок Δt, тоді

$$\tau = n * \Delta t \qquad \text{або} \qquad \Delta t = \tau / n.$$

Очевидно, що імовірність появи на елементарній ділянці Δt однієї події p визначиться в такий спосіб:

$$p=\lambda*\Delta t.$$

Будемо розглядати n ділянок як n дослідів, у кожному з яких подія може з'явитися з імовірністю $p=\lambda*\Delta t$ і не з'явитися з імовірністю $q=1-\lambda*\Delta t$.. Тоді імовірність, що в n дослідах подія з'явиться рівно k разів (тобто що на ділянку τ потрапить рівно k подій) за формулою Бернуллі визначиться в такий спосіб:

$$P_n(k) = C_n^k p^k q^{n-k} = C_n^k(\lambda\Delta t)^k(1-\lambda\Delta t)^{n-k} = \frac{n!}{k!(n-k)!}\frac{(\lambda\tau)^k}{n^k}\left(1-\lambda\frac{\tau}{n}\right)^{n-k} =$$

$$= \frac{1*2*...*(n-k)(n-k+1)...(n-1)*n}{1*2*...*(n-k)*k!*n^k}(\lambda\tau)^k\frac{(1-\lambda\frac{\tau}{n})^n}{(1-\lambda\frac{\tau}{n})^k}.$$

Дістанемо границю цього виразу при $n \to \infty$:

$$\lim_{n\to\infty}\frac{(n-k+1)*...*(n-1)*n}{n^k}*\frac{(\lambda\tau)^k}{k!}*\frac{(1-\lambda\frac{\tau}{n})^n}{(1-\lambda\frac{\tau}{n})^k},$$

чисельник останнього дробу перетворимо до вигляду:

$(1-\lambda\frac{\tau}{n})^n = \left[(1-\lambda\frac{\tau}{n})^{\frac{n}{\lambda\tau}}\right]^{\lambda\tau}$, в границі він прагне до $e^{-\lambda\tau}$, тоді отримаємо

$$P(k) = \frac{(\lambda\tau)^k}{k!}e^{-\lambda\tau}.$$

Отримана формула виражає закон розподілу Пуассона (5.9). Таким чином, коли імовірність p появи події А в кожному окремому досліді мала, а число дослідів n велике, біноміальний закон розподілу дискретної випадкової величини може бути приблизно замінений законом Пуассона.

Математичне сподівання випадкової величини, що має розподіл Пуассона, - це середнє число подій, що потрапляють на ділянку часу довжиною τ:

$$m_x=\lambda\tau. \tag{5.10}$$

Таким чином, закон розподілу Пуассона визначається одним параметром $a=\lambda\tau$, що є одночасно математичним сподіванням і дисперсією випадкової величини X:

$$D_x=\lambda\tau \tag{5.11}$$

Розподіл Пуассона з параметром a= np можна приблизно застосовувати замість біноміального, коли число дослідів n дуже велике, а імовірність p дуже мала, тобто в кожному окремому досліді подія А з'являється вкрай рідко. Розподіл Пуассона часто використовують, коли мають справу з числом подій, які з'являються на проміжку часу. Наприклад, число дефектів на новій ділянці шосе довжиною 10 км, число місць витоку води на 100 км водопроводу, число поломок надійного технічного пристрою за певний період часу, наприклад за рік.

Приклад 5.2. На АТС надходять виклики з інтенсивністю λ=0,8 1/хв. Визначити імовірність того, що протягом 2 хвилин а) не надійде жодного виклику; б) надійде рівно один виклик; в) надійде хоча б один виклик.

Розв'язання. Визначимо математичне сподівання числа викликів, що надходять на АТС, яке відповідає інтервалу часу t=2 хвилини:

$$a = \lambda * t = 0{,}8*2=1{,}6.$$

За формулою Пуассона імовірність подій визначиться в такий спосіб:

а) імовірність того, що протягом 2 хвилин не надійде жодного виклику

$$P(A) = P(0) = \frac{1{,}6^0}{0!} e^{-1{,}6} = 0{,}202;$$

б) імовірність того, що протягом 2 хвилин надійде рівно один виклик

$$P(B) = P(1) = \frac{1{,}6^1}{1!} e^{-1{,}6} = 1{,}6 * 0{,}202 = 0{,}323;$$

в) імовірність того, що протягом 2 хвилин надійде хоча б один виклик простіше визначити, використовуючи імовірність протилежної події:

$$P(C) = 1 - P(0) = 1 - 0{,}202 = 0{,}798.$$

Приклад 5.3. Потік вантажних поїздів, які прибувають на сортувальну станцію, має інтенсивність λ=4 поїзди в годину. Визначити імовірність того, що протягом 30 хвилин на сортувальну станцію прибуде: а) рівно один поїзд; б) хоча б один поїзд; в) не менше трьох поїздів.

Розв'язання. Визначимо середнє число поїздів, які прибувають на сортувальну станцію протягом 30 хвилин:

$$a = \lambda * t = 4*0{,}5=2.$$

Тоді за формулою Пуассона імовірність подій визначиться в такий спосіб:

а) імовірність того, що за 30 хвилин на сортувальну станцію прибуде рівно один поїзд

$$P(A) = P(1) = \frac{2^1}{1!} e^{-2} = 0{,}27;$$

б) імовірність того, що за 30 хвилин на сортувальну станцію прибуде хоча б один поїзд, простіше визначити, використовуючи імовірність протилежної події, яка полягає в тому, що не прибуде жодного поїзда:

$$P(0) = \frac{2^0}{0!} e^{-2} = 0{,}135,$$

тоді $$P(B) = 1 - P(0) = 1 - 0{,}135 = 0{,}865;$$

в) для визначення імовірності того, що за 30 хвилин на сортувальну станцію прибуде не менше трьох поїздів, знову скористаємося імовірністю протилежної події «менше трьох поїздів», що означає 0, або 1, або 2 поїзди. Імовірності того, що не прибуде жодного поїзда і що прибуде рівно один поїзд, нами вже визначені. Визначимо імовірність, що за 30 хвилин прибудуть рівно два поїзди:

$$P(2) = \frac{2^2}{2!} e^{-2} = 0{,}27.$$

Тоді імовірність події C визначиться в такий спосіб:

$$P(C) = 1 - P(0) - (1) - P(2) = 1 - 0{,}135 - 0{,}27 - 0{,}27 = 0{,}325.$$

Приклад 5.4. Космічні частки, які потрапляють в супутник, утворюють поле з щільністю λ=1 частка/м^2. Агрегат супутника, який знаходиться в полі часток, займає площу s=10 см^2. Для виходу з ладу агрегату свідомо досить влучення в нього двох часток. При влученні однієї частки він виходить із ладу з імовірністю p=0,5. Визначити імовірність виходу з ладу агрегату.

Розв'язання. Позначимо подію, що цікавить нас, А={вихід агрегату з ладу}. Цій події відповідають дві гіпотези:

H_1 ={в агрегат потрапила одна частка},

H_2 ={в агрегат потрапили дві частки}.

Умовні імовірності події А: $P(A/H_1)=0,5$, $P(A/H_2)=1$.

Імовірності гіпотез визначимо за законом розподілу Пуассона, параметр якого a=1*0,001=0,001:

$$P(H_1) = P(1) = \frac{0,001^1}{1!}e^{-0,001} = 0,000999 \quad ,$$

$$P(H_2) = 1 - P(0) - P(1) = 1 - \frac{0,001^0}{0!}e^{-0,001} - P(1) = 1 - 0,999 - 0,000999 = 10^{-6} \ .$$

За формулою повної імовірності дістанемо імовірність події А:

$$P(A)=P(H_1)P(A/H_1)+ P(H_2)P(A/H_2) = 0,000999*0,5+10^{-6}*1=5,005*10^{-4}.$$

Експонентний закон розподілу

Розглянемо потік однорідних подій, які ідуть одна за іншою у випадкові моменти часу. Його можна розглядати як послідовність точок на числовій осі, показану на рисунку 5.2.

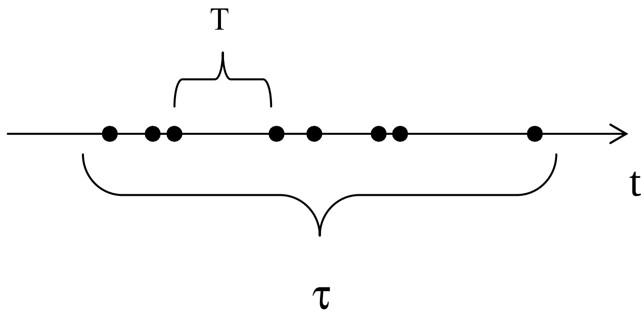


Рис. 5.2.

Якщо середнє число подій λ на ділянці часу довжиною τ залежить тільки від довжини ділянки, тобто постійне, не залежить від числа подій, що потрапляють на інші ділянки, і події в потоці йдуть по однієї, то такий потік називається **найпростішим**, або **стаціонарним пуассонівським потоком**.

Потік подій характеризується двома випадковими величинами: X - число подій, що потрапляють на інтервал часу τ, і T - проміжок часу між двома випадковими подіями в потоці. Зазначимо, що X є дискретною випадковою величиною, а T - безперервною.

Число подій, які потрапляють на ділянку часу τ в такому потоці, має закон розподілу Пуассона (5.9), де $a = \lambda * \tau$ - математичне сподівання числа подій;

λ - інтенсивність потоку подій (число подій в одиницю часу); τ - довжина ділянки часу.

Імовірність того, що на ділянці часу τ не відбудеться жодної події за законом Пуассона:

$$P(0) = \frac{a^0}{0!}e^{-a} = e^{-\lambda*\tau}. \tag{5.12}$$

Отримана імовірність стосовно випадкової величини проміжку часу між двома сусідніми подіями T - це імовірність того, що він опиниться більше деякого поточного t: $P\{T \geq t\}$. Тоді імовірність того, що T виявиться менше деякого поточного t ($P\{T < t\}$), є функцією розподілу T, і визначиться як імовірність протилежної події:

$$F(t) = P\{T < t\} = 1 - P\{T \geq t\} = 1 - e^{-\lambda t}, \tag{5.13}$$

Щільність розподілу T знайдемо як похідну функції розподілу F(t):

$$f(t) = d(t)/dt = \lambda e^{-\lambda t}, \tag{5.14}$$

Такий закон розподілу називається **експонентним**.

Визначимо числові характеристики випадкової величини, розподіленої за експонентним законом:

$$m_t = \int_{-\infty}^{+\infty} tf(t)dt = \int_0^{\infty} t\lambda e^{-\lambda t}dt = \lambda\int_0^{\infty} te^{-\lambda t}dt.$$

Проводячи інтегрування вроздріб, враховуємо, що при $t \to \infty$ $\exp\{-t\}$ прагне до нуля швидше, ніж зростає будь-яка степінь t, дістанемо:

$$m_x = 1/\lambda. \tag{5.15}$$

Таким чином, математичне сподівання випадкової величини, розподіленої за експонентним законом, зворотне параметру розподілу λ.

Визначимо дисперсію за формулою

$$D_t = \alpha_2 - m_t^2 = \lambda\int_0^{\infty} t^2 e^{-\lambda t}dt - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}, \tag{5.16}$$

звідки середнє квадратичне відхилення:

$$\sigma_x = 1/\lambda, \tag{5.17}$$

де λ - параметр розподілу (середнє число подій в одиницю часу).

Знайдемо імовірність влучення випадкової величини, що має експонентний розподіл, в інтервал значень (α, β). Нагадаємо, що ця імовірність дорівнює прирісту функції розподілу на інтервалі (α, β). З формули (5.13) маємо $F(\alpha) = 1 - e^{-\alpha t}$, $F(\beta) = 1 - e^{-\beta t}$, тоді

$$P\{\alpha \leq t \leq \beta\} = F(\beta) - F(\alpha) = 1 - e^{-\lambda\beta} - 1 + e^{-\lambda\alpha} = e^{-\lambda\alpha} - e^{-\lambda\beta}. \tag{5.18}$$

Експонентний розподіл часто використовується в теорії масового обслуговування, а також у теорії надійності.

Під надійністю розуміють імовірність безвідмовної роботи об'єкта протягом певного часу. Таким чином, надійність характеризується випадковою величиною T - час безвідмовної роботи. Функція розподілу часу безвідмовної роботи T - це, за визначенням, імовірність того, що T прийме значення, менше деякого поточного t:

$$F(t) = P\{T<t\}.$$

Очевидно, що F(t) є імовірністю того, що за час t наступить хоча б одна відмова. Імовірність безвідмовної роботи об'єкта протягом певного часу t (тобто імовірність, що відмова не наступить) є імовірністю протилежної події і визначиться в такий спосіб:

$$P(t) = P\{T>t\}.$$

На практиці час безвідмовної роботи T найчастіше розподілений за експонентним законом і його функція розподілу має вигляд

$$F(t) = 1 - e^{-\lambda t},$$

а імовірність безвідмовної роботи визначається в такий спосіб:

$$P(t) = 1 - F(t) = 1 - 1 + e^{-\lambda t} = e^{-\lambda t} \qquad (5.19)$$

Формула (5.19) дозволяє визначити імовірність безвідмовної роботи об'єкта протягом часу t.

Приклад 5.5. Є випадкова величина X з експонентним законом розподілу. Параметр розподілу λ=0,4. Визначити числові характеристики і функцію розподілу випадкової величини X, а також імовірність того, що вона прийме значення в інтервалі (6, 10).

Розв'язання. Числові характеристики випадкової величини X визначимо за формулами (5.15)-(5.17):

$m_x = 1/\lambda = 1/0,4 = 2,5$; $D_x = 1/\lambda^2 = 1/(0,4)^2 = 6,25$; $\sigma_x = m_x = 2,5$.

Щільність розподілу :

$$f(x) = \begin{cases} 0, & при \quad x < 0 \\ 0,4e^{-0,4x}, & при \quad x \ge 0 \end{cases}.$$

Функція розподілу:

$$F(x) = \begin{cases} 0, & при \quad x < 0 \\ 1 - e^{-0,4x}, & при \quad x \ge 0 \end{cases}.$$

Визначимо імовірність того, що випадкова величина X прийме значення в інтервалі (6, 10), для чого скористаємося формулою (5.19):

$$P\{6 \le X \le 10\} = e^{-\lambda\alpha} - e^{-\lambda\beta} = e^{-0,4*6} - e^{-0,4*10} = 0,0907 - 0,0183 = 0,0724.$$

Приклад 5.6. Визначити імовірність безвідмовної роботи технічного пристрою протягом одного року, якщо інтенсивність відмов дорівнює 0,05 відмови на рік, а час безвідмовної роботи розподілений за експонентним законом.

Розв'язання. Скористаємося формулою (5.19), де λ= 0,05, t= 1 рік. Імовірність того, що протягом одного року технічний пристрій не відмовить жодного разу

$$P(t) = e^{-\lambda t} = e^{-0,05*1} = 0,951.$$

Нормальний закон розподілу імовірностей

Цей закон ще називається законом Гаусса, оскільки був запропонований ним при дослідженні помилок точних вимірювань (зазначимо, що помилки грубих вимірювань мають інший розподіл імовірностей). Закон базується на двох посилках:

1)помилки різного знака, однакові за розміром, рівноімовірні;

2)малі помилки більш імовірні, ніж великі (промахи).

Цим посилкам відповідає горбоподібна крива, симетрична щодо середнього значення помилки вимірювання (рис. 5.3.).

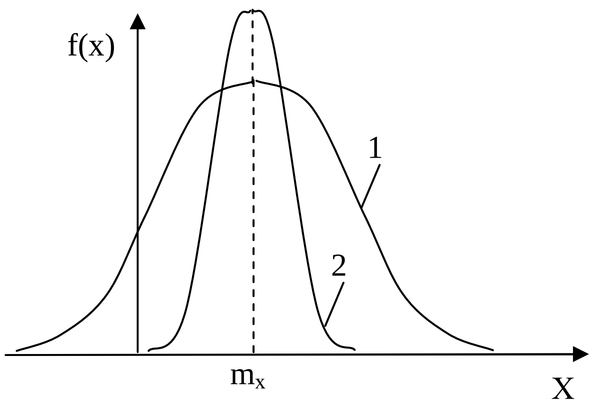


Рис.5.3.

Отримана крива апроксимується виразом

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}. \qquad (5.20)$$

Звідси видно, що нормальний закон розподілу визначається двома параметрами m_x і σ_x.

На практиці закон нормального розподілу зустрічається дуже часто, тому що існує велике число нормально розподілених випадкових величин. Якщо відомі параметри m_x і σ_x, то із сімейства всіх кривих нормального розподілу виділяють одну з певною щільністю.

При $x=m_x$ щільність розподілу максимальна і дорівнює

$$f(x)=\frac{1}{\sigma_x\sqrt{2\pi}}.$$

Оскільки інтеграл від щільності розподілу в нескінченних межах дорівнює одиниці, тобто крива на рис. 5.2 обмежує площу, яка дорівнює одиниці, то чим менше параметр σ_x, тим крутіше спадає крива і тим менше розкидані значення X на числовій осі.

Визначимо імовірність влучення нормально розподіленої випадкової величини в інтервал значень (α, β) як інтеграл від щільності розподілу в межах від α до β:

$$P\{\alpha\le X\le\beta\}=\int_\alpha^\beta f(x)dx=\int_\alpha^\beta\frac{1}{\sigma_x\sqrt{2\pi}}e^{-\frac{(x-m_x)^2}{2\sigma_x^2}}=\left|\begin{matrix}t=\dfrac{x-m_x}{\sigma_x}\\ dx=\sigma_x dt\end{matrix}\right|=\frac{1}{\sqrt{2\pi}}\int_{t_1}^{t_2}e^{-\frac{t^2}{2}}dt, \qquad (5.21)$$

де $t_1=\dfrac{\alpha-m_x}{\sigma_x}$; $t_2=\dfrac{\beta-m_x}{\sigma_x}$.

Оскільки інтеграл (5.21) не береться в елементарних функціях, для визначення імовірностей, пов'язаних з нормально розподіленою випадковою величиною, користуються функцією Лапласа (інтегралом імовірностей):

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{0}^{x}e^{-\frac{t^2}{2}}dt\ , \qquad (5.22)$$

значення якої наведені в довідкових таблицях.

Імовірність влучення випадкової величини X на ділянку значень (α, β) виражається через функцію Лапласа формулою

$$P\{\,\alpha < X < \beta\,\}=\Phi\left(\frac{\beta-m_x}{\sigma}\right)-\Phi\left(\frac{\alpha-m_x}{\sigma}\right). \qquad (5.23)$$

При обчисленні імовірностей користуються наступними властивостями функції Лапласа (рис. 5.4):

1) при x=0 Ф(x)=0;
2) при x=(Ф(x)=0,5;
3) при x=-(Ф(x)=-0,5;
4) функція Ф(x) є непарною функцією, тобто Ф(-x) =-Ф(x)

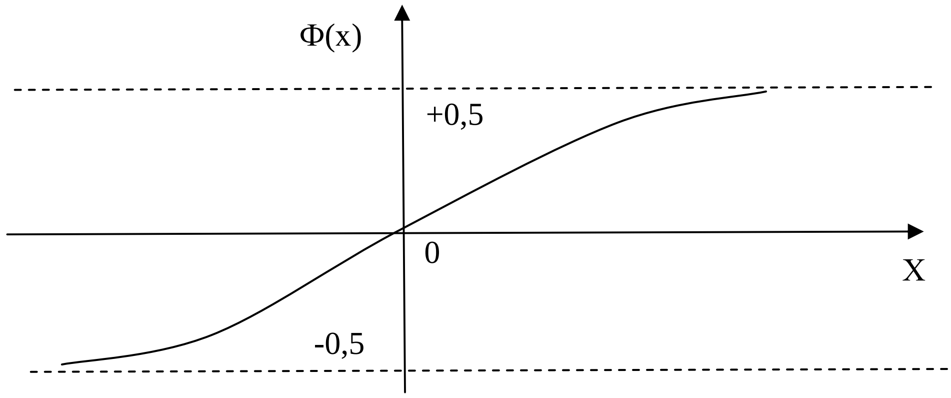


Рис. 5.4 - Крива функції Лапласа

Таким чином, всі можливі значення функції Лапласа належать інтервалу (-0,5;+0,5), причому при |x| > 4 можна вважати, що Ф(x) ≈±0,5.

Скористуємося формулою (5.23), і визначимо функцію розподілу для випадкової величини, розподіленої нормально:

$$F(x)=\int_{-\infty}^{x}f(x)dx=\Phi\left(\frac{x-m_x}{\sigma_x}\right)-\Phi\left(\frac{-\infty-m_x}{\sigma_x}\right)=\Phi\left(\frac{x-m_x}{\sigma_x}\right)+0{,}5\,. \qquad (5.24)$$

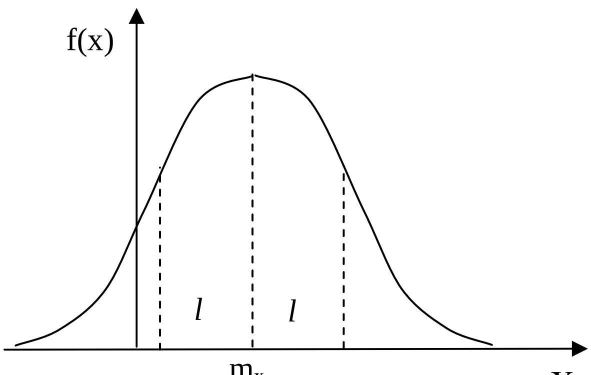


Рис.5.5 - Значення, симетричні відносно математичного сподівання

Оскільки нормальний розподіл є симетричним, звичайно становить інтерес імовірність влучення нормально розподіленої випадкової величини в ділянку значень, симетричну щодо її математичного сподівання (рис 5.5):

$$P\ (|x-m_x| < l\ )$$

Для її визначення скористаємося також формулою (5.23).

$$P\{|x-m_x| < l\} = P\{m_x - l < X < m_x + l\} =$$

$$= \Phi\left(\frac{m_x + l - m_x}{\sigma_x}\right) - \Phi\left(\frac{m_x - l - m_x}{\sigma_x}\right) = 2\Phi\left(\frac{l}{\sigma_x}\right).$$

Візьмемо тепер $l = \sigma_x$, тоді:

$$P\{\,|x-m_x| < \sigma\,\} = 2 * \Phi\left(\frac{\sigma_x}{\sigma_x}\right) = 2 * \Phi(1) = 0{,}68\,.$$

Таким чином, 68 % значень будь-якої нормально розподіленої випадкової величини лежать в інтервалі ($m_x \pm \sigma_x$).

Нехай $l = 2\sigma_x$, тоді

$$P\{\,|x-m_x| < 2\sigma\,\} = 2 * \Phi\left(\frac{2\sigma_x}{\sigma_x}\right) = 2 * \Phi(2) = 0{,}95\,.$$

Отже 95 % значень будь-якої нормально розподіленої випадкової величини лежать в інтервалі ($m_x \pm 2\sigma_x$).

Якщо $l = 3\sigma_x$, то

$$P\{\,|x-m_x| < 3\sigma\,\} = 2 * \Phi\left(\frac{3\sigma_x}{\sigma_x}\right) = 2 * \Phi(3) = 0{,}997\,.$$

Тобто 99,7 % значень будь-якої нормально розподіленої випадкової величини лежать в інтервалі ($m_x \pm 3\sigma_x$).

Ця властивість випадкових величин, розподілених нормально, називається «правилом трьох сигма».

Приклад 5.6. Відомі імовірнісні характеристики нормально розподіленої випадкової величини X: m =17; σ=0,6. Знайти імовірність події $P(\alpha<X<\beta)$; імовірність того, що $P(|x-m| < \delta)$, якщо α =16,8; β = 17,2; δ=0,3.

Розв'язання. Обчислимо імовірність, що X належить інтервалу (16,8; 17,2):

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - m}{\sigma}\right) - \Phi\left(\frac{\alpha - m}{\sigma}\right) =$$

$$= \Phi\left(\frac{17{,}2 - 17}{0{,}6}\right) - \Phi\left(\frac{16{,}8 - 17}{0{,}6}\right) = \Phi\left(\frac{1}{3}\right) - \Phi\left(-\frac{1}{3}\right) = 2 * \Phi\left(\frac{1}{3}\right) = 0{,}26.$$

Визначимо імовірність того, що X відхилиться від свого середнього значення m менш ніж на δ:

$$P(\,|x-17| < 0{,}3\,) = 2 * \Phi\left(\frac{0{,}3}{0{,}6}\right) = 0{,}38\,.$$

Приклад 5.7. Деталь, виготовлена автоматом, вважається придатною, якщо відхилення X контрольованого розміру від номіналу не перевищує 10 мм. Точність виготовлення деталей характеризується σ=0,5. Вважаючи, що X розподілена нормально, визначити, скільки відсотків придатних деталей виготовляє автомат.

Розв'язання. Іншими словами, треба визначити імовірність того, що помилка X потрапить у симетричний відносно m_x інтервал, який дорівнює 10 мм:

$$P\left\{\left|\overset{o}{X}\right| < 10\right\} = 2 * \Phi\left(\frac{l}{\sigma}\right) = 2 * \Phi\left(\frac{10}{5}\right) = 0{,}95\,.$$

Автомат випускає 95% придатних деталей.

Приклад 5.8. У нормально розподіленій сукупності 15% значень X менше 12 і 40% значень X більше 16,2. Знайти середнє значення і середнє квадратичне відхилення даного розподілу.

Розв'язання. З умови задачі випливає, що

$$P\{X<12\}=0{,}15 \text{ і } P\{X<16{,}2\}=0{,}6.$$

Скористуємось формулою (5.17) і запишемо систему рівнянь:

$$\begin{cases} \Phi\left(\dfrac{12-m_x}{\sigma_x}\right)+0{,}5=0{,}15 \\ \Phi\left(\dfrac{16{,}2-m_x}{\sigma_x}\right)+0{,}5=0{,}6 \end{cases} \Rightarrow \begin{cases} \Phi\left(\dfrac{12-m_x}{\sigma_x}\right)=-0{,}35 \\ \Phi\left(\dfrac{16{,}2-m_x}{\sigma_x}\right)=0{,}1 \end{cases} \Rightarrow \begin{cases} \dfrac{12-m_x}{\sigma_x}=-1{,}04 \\ \dfrac{16{,}2-m_x}{\sigma_x}=0{,}25 \end{cases} \Rightarrow .$$

$$\Rightarrow \begin{cases} m_x = 15{,}386 \\ \sigma_x = 3{,}256 \end{cases}.$$

Параметри розподілу: m_x = 15,386, σ_x =3,256.

Приклад 5.9. Коробки з шоколадом упаковує автомат. Їхня середня маса 1,06 кг. Відомо, що 5% коробок мають масу менше 1 кг. Визначити відсоток коробок, маса яких перевищує 940 г?

Розв'язання. Визначимо відсоток коробок, маса яких менше 940 г, тоді виявиться, що інші коробки мають масу більше 940 г. Скористаємося формулою (5.22):

$$P\{X < 0{,}940\} = F(0{,}940) = \Phi\left(\frac{0{,}94 - m_x}{\sigma_x}\right) + 0{,}5\,.$$

Математичне сподівання відомо з умови задачі m_x = 1,06, а для визначення невідомого середнього квадратичного відхилення скористаємося тим, що за умовою 5% коробок мають масу менше 1 кг:

$$P\{X < 1{,}0\} = F(1{,}0) = \Phi\left(\frac{1{,}0 - 1{,}06}{\sigma_x}\right) + 0{,}5 = 0{,}05\,,$$

звідки маємо:

$$\Phi\left(\frac{1{,}0-1{,}06}{\sigma_x}\right) = -0{,}45\,, \qquad \frac{1{,}0-1{,}06}{\sigma_x} = -1{,}655\,, \qquad \sigma_x = 0{,}03625.$$

Тоді

$$P\{X < 0{,}940\} = \Phi\left(\frac{0{,}94 - 1{,}06}{0{,}03625}\right) + 0{,}5 = \Phi(-3{,}31) + 0{,}5 = -0{,}499 + 0{,}5 = 0{,}001\,.$$

Відсоток коробок, маса яких перевищує 940 г, становить 99,9%.

Поняття про центральну граничну теорему

Центральна гранична теорема - це загальна назва групи теорем, що стосуються законів розподілу суми випадкових величин, зміст яких зводиться до наступного:

закон розподілу суми незалежних випадкових величин наближається до нормального закону розподілу при необмеженому збільшенні числа доданків, якщо всі величини мають скінченні математичні сподівання і дисперсії і жодна з величин за значенням різко не відрізняється від інших.

Познайомимося з формулюванням двох теорем.

Теорема 1. Якщо незалежні $X_1, X_2, \ldots X_n$ мають той самий закон розподілу з математичним сподіванням m і дисперсією D, то при необмеженому збільшенні n закон розподілу суми $X_1 + X_2 + \ldots + X_n$ необмежено наближається до нормального.

Теорема 2 (центральна гранична теорема Ляпунова). Якщо випадкова величина $Y = X_1 + X_2 + \ldots + X_n$, де вплив кожного з доданків на всю суму рівномірно малий, то величина Y має розподіл, близький до нормального, і тим ближче, чим більше n.

Дослід показує, що закон розподілу суми незалежних випадкових величин, порівнянних за своїм розсіюванням, досить швидко наближається до нормального. Уже при числі доданків порядку десяти закон розподілу суми можна замінити нормальним. При цьому важливо те, що закони розподілу доданків можуть бути будь-якими, заздалегідь невідомими.

Багато випадкових величин можна розглядати як суму незалежних доданків, наприклад, помилки вимірювань, відхилення розмірів деталей, число продажів деякого товару, обсяг прибутку від реалізації, валютні курси і т.д.

Якщо є доданки X_i, що мають переважний вплив на величину Y, то робити твердження про нормальний розподіл Y не можна. У цьому разі закон розподілу Y буде визначатися композицією законів розподілу доданків, вплив яких на Y великий.

Закон рівномірної щільності

Безперервна випадкова величина X має рівномірний розподіл на ділянці від α до β, якщо її щільність розподілу на цій ділянці постійна:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & при \quad x \in (a,b); \\ 0 & при \quad x \notin (a,b) \end{cases}. \qquad (5.25)$$

Наприклад, помилка X при грубому вимірюванні, що може приймати з постійною щільністю імовірності будь-яке значення між двома сусідніми цілими діленнями. Грубий вимір відрізняється від точного тим, що результат грубих вимірювань при повторенні завжди той самий, при точному ж вимірюванні результат щоразу змінюється. Іншим прикладом рівномірного розподілу є абсциса навмання поставленої точки на відрізку [a, b].

Визначимо числові характеристики випадкової величини, розподіленої рівномірно.

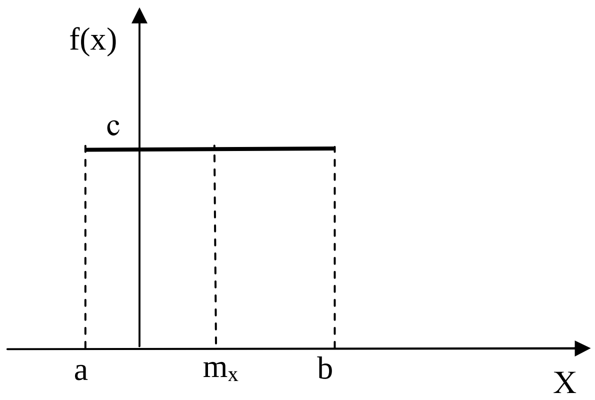


Рис.5.6.

Математичне сподівання:

$$m_x = \int_{-\infty}^{+\infty} x f(x)dx = \int_a^b x\frac{1}{b-a}dx = \frac{1}{b-a} * \frac{x^2}{2}\Bigg|_a^b = $$

$$= \frac{b+a}{2} \tag{5.26}$$

Дисперсія:

$$D_x = \int_{-\infty}^{+\infty}(x-m_x)^2 f(x)dx = \int_a^b (x - \frac{b+a}{2})^2 \frac{1}{b-a}dx =$$

$$= \frac{(b-a)^2}{12} \tag{5.27}$$

Середнє квадратичне відхилення:

$$\sigma_x = \sqrt{D_x} = \frac{b-a}{2\sqrt{3}}. \tag{5.28}$$

Визначимо імовірність влучення значень рівномірно розподіленої випадкової величини на інтервал (α, β):

$$P\{\alpha < X < \beta\} = \int_\alpha^\beta \frac{1}{b-a}dx = \frac{\beta-\alpha}{b-a} \tag{5.29}$$

Функція рівномірного розподілу:

$$F(x) = \int_\alpha^x \frac{1}{b-a}dx = \frac{x-\alpha}{b-a}. \tag{5.30}$$

Приклад 5.10. Довжину кімнати виміряють рулеткою з грубими поділками (10 см). Округлення проводять до найближчого цілого. X - помилка вимірювання. Знайти її щільність розподілу, функцію розподілу і числові характеристики.

Розв'язання. Довжина кімнати L з урахуванням помилки визначиться як L±5 см, тобто випадкова величина X змінюється в межах -5<X<+5. Оскільки крива щільності розподілу обмежує площу, рівну одиниці, значення f(x) дістанемо в такий спосіб:

$$f(x) = \frac{1}{b-a} = \frac{1}{5-(-5)} = 0{,}1.$$

Математичне сподівання визначимо за формулою (5.26)

$$m_x = \frac{a+b}{2} = \frac{-5+5}{2} = 0.$$

Дисперсію визначимо за формулою (5.27):

$$D_x = \frac{(b-a)^2}{12} = \frac{(5+5)^2}{12} = 8{,}33; \qquad \sigma x = \sqrt{8{,}33} = 2{,}89.$$

Приклад 5.11. Поїзди метро йдуть регулярно з інтервалом 2 хвилини. Пасажир виходить на платформу у випадковий момент часу, не пов'язаний з розкладом поїздів. Випадкова величина T - час очікування поїзда. Знайти: а) щільність розподілу і числові характеристики випадкової величини T; б) імовірність того, що чекати доведеться не більше 0,5 хвилини.

Розв'язання. Оскільки поїзди під'їжджають до станції рівномірно, закон розподілу випадкової величини T – рівномірний, тобто $f(x) = \frac{1}{b-a}$, де (b-a) – інтервал руху поїздів, причому a =0, b =2, тоді:

а) $f(x) = \frac{1}{2-0} = 0{,}5$; $m_x = \frac{b+a}{2} = \frac{2+0}{2} = 1$ хв.; $D_x = \frac{(b-a)^2}{12} = \frac{(2-0)^2}{12} = \frac{1}{3}$;

б) імовірність того, що чекати доведеться не більше 0,5 хвилини, визначимо за формулою (5.30):

$$P\{T < 0{,}5\} = F(x) = \frac{0{,}5-0}{2-0} = 0{,}25\,.$$

Запитання для самоперевірки:

1. Яким умовам повинні задовольняти повторні незалежні випробування?
2. Як визначають числові характеристики випадкової величини, розподіленої за законом Бернуллі?
3. Який зв'язок існує між біноміальним і пуассонівським розподілами?
4. Яким умовам повинна задовольняти випадкова величина, підпорядкована закону Пуассона?
5. Як визначають числові характеристики закону розподілу Пуассона?
6. Якими параметрами визначається експонентний закон розподілу випадкової величини?
7. Чому дорівнює щільність імовірності випадкової величини з нормальним законом розподілу?
8. Якими параметрами визначається нормальний закон розподілу випадкової величини?
9. Як змінюється графік нормального закону із зміною середнього квадратичного відхилення випадкової величини?
10. Як визначити імовірність влучення нормально розподіленої випадкової величини на задану ділянку?
11. Поясніть імовірнісний смисл параметрів нормального розподілу.
12. Поясніть смисл центральної граничної теореми.

Задачі для самостійного розв'язання

5.1. Випадкова величина X підлегла закону розподілу Пуассона з математичним сподіванням *a*=3. Побудувати многокутник розподілу і функцію розподілу випадкової величини X. Знайти імовірність того, що випадкова величина X прийме значення менше, ніж її математичне сподівання.

5.2. Випадкова величина X підпорядкована експонентному закону розподілу з параметром μ:

$$f(x)=\begin{cases}\mu e^{-\mu x} & при\ x>0\\ 0 & при\ x<0\end{cases}$$

Побудувати криву розподілу, визначити функцію розподілу і знайти імовірність того, що випадкова величина X прийме значення, менше її математичного сподівання.

5.3. Вимірювальний прибор має систематичну помилку 5 м і середню квадратичну помилку 75 см. Яка імовірність того, що помилка вимірювання не перевершить за абсолютною величиною 5 м?

5.4. Випадкова величина X підпорядкована нормальному закону з математичним сподіванням, рівним нулю. Імовірність влучення цієї випадкової величини на ділянку від -α до +α дорівнює 0,5. Знайти середнє квадратичне відхилення і написати вираз нормального закону.

5.5. У світлофорі на перехресті 1 хвилину горить зелене світло і 0,5 хвилини червоне. Автомобіль під'їжджає до перехрестя у випадковий момент, не пов'язаний з роботою світлофора. Знайти імовірність того, що він проїде перехрестя, не зупиняючись.

Тема 6. СИСТЕМА ВИПАДКОВИХ ВЕЛИЧИН. ЗАКОНИ РОЗПОДІЛУ і ЧИСЛОВІ ХАРАКТЕРИСТИКИ СИСТЕМИ

Багатовимірна випадкова величина.

При вивченні випадкових явищ іноді доводиться використовувати дві, три і більше випадкових величин. Спільний розгляд двох або декількох випадкових величин приводить до поняття системи випадкових величин. Систему декількох випадкових величин X, Y,... , W позначають (X,Y,... ,W) і називають **багатомірною випадковою величиною**. При вивченні багатовимірної випадкової величини недостатньо вивчити окремо складаючи її випадкові величини, а необхідно враховувати також зв'язки між цими величинами.

Найбільше практичне значення має система двох випадкових величин. Для характеристики системи двох випадкових величин використовують закони розподілу системи і числові характеристики системи. Найбільш простим є закон розподілу системи двох дискретних випадкових величин, що уявляє собою кореляційну таблицю (кореляція від англійського correlation — співвідношение, відповідність), в якій перший рядок містить всі значення випадкової величини X, а перший стовпець - всі значення випадкової величини B. В ij-й клітині таблиці записують імовірність події $\{X=x_i, Y=y_j\}$.

x_i / y_j	x_1	x_2	…	x_n
y_1	$p\{x_1, y_1\}$	$p\{x_2, y_1\}$	…	$p\{x_n, y_1\}$
y_2	$p\{x_1, y_2\}$	$p\{x_2, y_2\}$	…	$p\{x_n, y_2\}$
…	…	…	…	…
y_m	$p\{x_1, y_m\}$	$p\{x_2, y_m\}$	…	$p\{x_n, y_m\}$

Сума всіх імовірностей у кореляційній таблиці дорівнює одиниці

$$\sum_{i=1}^{n} \sum_{j=1}^{m} p\{x_i, y_j\} = 1.$$

Функція розподілу системи двох випадкових величин

Функція розподілу системи двох випадкових величин (X,Y) дорівнює імовірності того, що випадкова величина X прийме значення, менше x і випадкова величина Y прийме значення, менше y:

$$F(x,y) = P\{X < x, Y < y\}. \tag{6.1}$$

Геометрична інтерпретація функції розподілу є імовірність влучення X і Y у нескінченний квадрат з координатами вершини (x,y) (рис. 6.1).

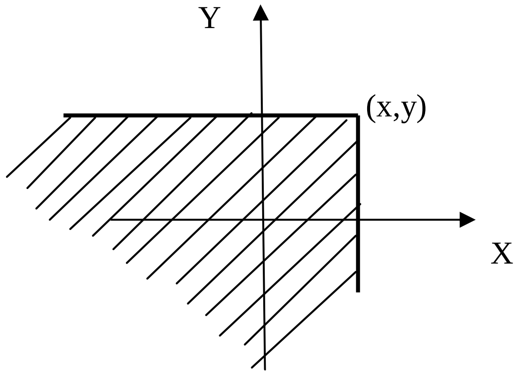


Рис. 6.1 - Геометрична інтерпретація функції розподілу

Властивості функції розподілу системи:

1. Функція розподілу - неубутна функція, тобто

$$F(x_2,y) \geq F(x_1,y), \quad \text{якщо} \quad x_2 > x_1 .$$
$$F(x,y_2) \geq F(x,y_1), \quad \text{якщо} \quad y_2 > y_1 .$$

2. Функція розподілу дорівнює нулю, якщо хоча б один з аргументів обертається в мінус нескінченність:

$$F(x, -\infty) = F(-\infty, y) = F(-\infty, -\infty) = 0.$$

3. Функція розподілу, якщо хоча б один з аргументів обертається в плюс нескінченність, дорівнює функції розподілу компонента системи, що залишився

$$F(x, +\infty) = F(x);$$
$$F(+\infty, y) = F(y);$$
$$F(+\infty, +\infty) = 1.$$

Щільність розподілу

Щільність розподілу імовірностей системи двох випадкових величин (X,Y) f(x,y) являє собою другу змішану похідну від F(x,y):

$$f(x,y) = \frac{\partial^2 F(x,y)}{\partial x \partial y}. \tag{6.2}$$

Тут припускається, що F(x,y) безперервна й двічі диференційована. Властивості щільності розподілу імовірностей:

1. Щільність розподілу невід’ємна, тобто $f(x,y) \geq 0$;

2. Функція розподілу системи двох випадкових величин (X,Y) дорівнює подвійному інтегралу від щільності розподілу системи:

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(x,y)dxdy,$$

де $f(x,y)dxdy$ - елемент імовірності, що являє собою імовірність влучення системи в елементарний прямокутник dxdy і дорівнює об'єму паралелепіпеда f(x)dxdy.

3. Подвійний інтеграл від щільності розподілу в нескінченних межах дорівнює одиниці:

$$\iint\limits_{-\infty}^{+\infty} f(x,y)dxdy = 1.$$

Числові характеристики системи випадкових величин

За аналогією з числовими характеристиками однієї випадкової величини числовими характеристиками системи двох випадкових величин є початкові й центральні моменти α_{ks} і μ_{ks}, причому порядок моменту визначається сумою індексів k+s.

Початковим моментом порядку k,s системи двох випадкових величин (X, Y) називається математичне сподівання добутку випадкової величини X в степені k на випадкову величину Y в степені s:

$$\alpha_{k,s} = M[X^k Y^s]. \tag{6.3}$$

Центральним моментом порядку k,s системи випадкових величин (X, Y) називається математичне сподівання добутку центрованої величини X в степені k на центровану випадкову величину Y в степені s

$$\mu_{k,s} = M[\overset{0}{X}{}^k \overset{0}{Y}{}^s] \tag{6.4}$$

Зокрема початковими моментами першого порядку є математичні сподівання випадкових компонентів системи X і Y:

$$\alpha_{1,0} = M[X^1 Y^0] = M[X];$$

$$\alpha_{0,1} = M[X^0 Y^1] = M[Y].$$

Центральними моментами другого порядку є дисперсії випадкових компонентів системи X і Y:

$$\mu_{2,0} = M[\overset{0}{X}{}^2 \overset{0}{Y}{}^0] = M[\overset{0}{X}{}^2] = D_x;$$

$$\mu_{0,2} = M[\overset{0}{X}{}^0 \overset{0}{Y}{}^2] = M[\overset{0}{Y}{}^2] = D_y.$$

Для опису системи двох випадкових величин крім математичних сподівань і дисперсій X і Y використовують кореляційний момент і коефіцієнт кореляції. Кореляційним моментом є другий змішаний центральний момент:

$$\mu_{x,y} = M[\overset{0}{X}\overset{0}{Y}] = K_{xy}. \tag{6.5}$$

Для дискретної випадкової величини K_{xy} визначають за формулою

$$K_{x,y} = \sum_{i=1}^{n}\sum_{j=1}^{m} \overset{0}{x}_i \overset{0}{y}_j p_{ij} = \sum_{i=1}^{n}\sum_{j=1}^{m}(x_i - m_x)(y_j - m_y)p_{ij}, \tag{6.6}$$

де $p_{ij} = P\{X=x_i|Y=y_j\}$ – умовна імовірність, тобто імовірність того, що X прийме значення x_i за умови, що Y прийме значення y_j.

Для безперервної випадкової величини

$$K_{xy} = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty}(x - m_x)(y - m_y)f(x,y)dxdy. \tag{6.7}$$

Якщо події $P\{X=x_i\}$ і $P\{Y=y_j\}$ незалежні, то імовірність їхньої спільної появи за теоремою множення дорівнює

$$p_{ij} = P\{X=x_i\} * P\{Y=y_j\} = p_i * p_j.$$

Тоді для кореляційного моменту справедливий вираз

$$K_{x,y} = \sum_{i=1}^{n} \overset{0}{x_i}\, p_i \sum_{j=1}^{m} \overset{0}{y_j}\, p_j = 0\,,$$

оскільки співмножники є центральними моментами першого порядку випадкових величин X і Y, а тому дорівнюють нулю. Таким чином, кореляційний момент є характеристикою зв'язку між величинами X і Y, у випадку незалежних X і Y він дорівнює нулю.

Як другий змішаний центральний момент кореляційний момент містить також розсіювання випадкових величин X і Y відносно ода одної. Тому він не може характеризувати тісноту зв'язку між X і Y. Для визначення тісноти зв'язку між X і Y використовують коефіцієнт кореляції r_{xy}, що визначається за формулою

$$r_{xy} = \frac{K_{xy}}{\sigma_x \sigma_y}. \tag{6.8}$$

Переконаємося в тому, що r_{xy} характеризує ступінь тісноти лінійного зв'язку між двома випадковими величинами. Нехай випадкова величина Y функціонально (жорстко) залежить від випадкової величини X, причому залежність ця лінійна:

$$Y = a+b.$$

Визначимо математичне сподівання

$$M[Y] = M[a+b] = \Sigma(ax_i+b)p_i = a\Sigma x_i p_i + \Sigma b p_i = a[X]+b.$$

Знайдемо дисперсію:

$$D[Y] = M[\overset{o}{Y}{}^2] = M[(Y-m_y)^2] = M[Y^2-2my+m_y^2] =$$
$$=M[(a+b)^2-2(am_x+b)(a+b)+(am_x+b)^2] = M[a^2(X-m_x)^2]= a^2D_x,$$

і середнє квадратичне відхилення Y $\sigma_y = |a|\sigma_x$.

Визначимо коефіцієнт кореляції для жорстко зв'язаних X і Y, для чого виразимо $\overset{o}{Y}$ через $\overset{o}{X}$: $\overset{o}{Y} = Y - m_y = aX + b - am_x - b = a(X - m_x) = a\overset{o}{X}$, тоді кореляційний момент дорівнюватиме

$$K_{xy} = M[\overset{o}{Y}\overset{o}{X}] = [a\overset{o}{X}\overset{o}{X}] = aD_x$$

а коефіцієнт кореляції

$$r_{xy} = \frac{K_{xy}}{\sigma_x\sigma_y} = \frac{aD_x}{\sigma_x|a|\sigma_x} = \frac{a\sigma_x^2}{\sigma_x|a|\sigma_x} = \frac{a}{|a|}.$$

Таким чином, якщо зв'язок функціональний, то коефіцієнт кореляції дорівнює 1, причому

$$r_{xy} = \begin{cases} -1 \ \text{при} \ \ a<0 \\ +1 \ \text{при} \ \ a>0 \end{cases} \tag{6.9}$$

У загальному випадку коефіцієнт кореляції лежить у межах $-1 \le r_{xy} \le +1$ і дорівнює нулю, якщо X і Y незалежні.

Дві випадкові величини X і Y називаються корельованими, якщо їхній кореляційний момент не дорівнює нулю. Дві корельовані випадкові величини обов'язково залежні. Якщо кореляційний момент випадкових величин X і Y дорівнює нулю, то ці випадкові величини некорельовані, але вони можуть опинитись залежними. Таким чином, поняття **корельованості** і **залежності** двох випадкових величин різні.

Приклад 6.1. Матриця розподілу випадкового вектора (X, Y) має вигляд

x_i \ y_i	0	2	5
1	0,1	0	0,2
2	0	0,3	0
4	0,1	0,3	0

Знайти числові характеристики.

Розв'язання. Для визначення числових характеристик випадкової величини X побудуємо її ряд розподілу:

x_i	1	2	4
p_i	0,3	0,3	0,4

де P{X=1}=0,1+0+0,2=0,3; P{X=2}=0+0,3+0=0,3; P{X=4}=0,1+0,3+0=0,4.

Тоді $m_x = \Sigma x_i p_i = 1*0,3+2*0,3+4*0,4=2,5$;

$D_x = \Sigma(x_i - m_x)^2 p_i =(1-2,5)^2*0,3+(2-2,5)^2*0,3+(4-2,5)^2*0,4=1,65$;

$\sigma_x = \sqrt{D_x} = \sqrt{1,65} = 1,285$.

Аналогічно, побудуємо ряд розподілу випадкової величини Y:

y_i	0	2	5
p_i	0,2	0,6	0,2

і визначимо її числові характеристики:

$m_y = \Sigma y_i p_i = 0*0,2+2*0,6+5*0,2=2,2$;

$D_y = \Sigma(y_i - m_y)^2 p_i =(0-2,2)^2*0,3+(2-2,2)^2*0,6+(5-2,2)^2*0,2=2,54$;

$\sigma_y = \sqrt{D_y} = \sqrt{2,54} = 1,6$.

Визначимо кореляційний момент системи випадкових величин X, Y:

$$K_{xy} = \sum_{i=1}^{3}\sum_{j=1}^{3}(x_i - m_x)(y_j - m_y)p_{ij} = (1-2,5)\left[(0-2,2)*0,1+(2-2,2)*0+(5-2,2)*0,2\right]+$$

$$+(2-2{,}5)\big[(0-2{,}2)*0+(2-2{,}2)*0{,}3+(5-2{,}2)*0\big]+$$
$$+(4-2{,}5)\big[(0-2{,}2)*0{,}1+(2-2{,}2)*0{,}3+(5-2{,}2)*0\big]=-0{,}9.$$

Знайдемо коефіцієнт кореляції:

$$r_{xy}=\frac{K_{xy}}{\sigma_x\sigma_y}=\frac{-0{,}9}{1{,}285*1{,}6}=-0{,}438.$$

Приклад 6.2. Двічі кидають гральну кістку. Випадкові величини X - число появ шістки, Y - число появ парної цифри. Описати закони розподілу випадкових величин X і Y; описати закон розподілу випадкового вектора (X, Y); установити, залежні X і Y, або незалежні.

Розв'язання. Знайдемо закон розподілу випадкового вектора (X, Y), склавши матрицю розподілу, оскільки X і Y - дискретні.

x_i \ y_i	0	1	2	$P\{X=x_i\}$
0	1/4	1/3	1/9	25/36
1	0	1/6	1/9	10/36
2	0	0	1/36	1/36
$P\{Y=y_i\}$	1/4	1/2	1/4	1

Для визначення умовних імовірностей $P_{ij}=P\{X=x_i/Y=y_j\}$ скористуємося теоремами додавання і множення імовірностей.

Дістанемо ряд розподілу випадкової величини X, для чого підсумовуємо умовні імовірності в кореляційній таблиці за рядками і Y, для чого підсумовуємо умовні імовірності за стовпцями.

Установимо, залежні чи незалежні X і Y. Відомо, що у випадку незалежних подій умовна імовірність події дорівнює його безумовної імовірності, але рівність $P\{y_j/x_i\}=P\{y_j\}$ не дотримується. Наприклад,

$$P_{22}\{y_j=1,\ x_i=1\}=P\{x_i=1\}*P\{y_j=1/x_i=1\}=1/6.$$

Таким чином, X і Y залежні.

Приклад 6.3. Для умов попереднього прикладу знайти числові характеристики випадкового вектора (X, Y).

Розв'язання. Визначимо математичні сподівання X і Y:

$$m_x=\Sigma x_i p_i=0*\frac{25}{36}+1*\frac{5}{18}+2*\frac{1}{36}=\frac{1}{3};$$

$$m_y=\Sigma y_i p_i=0*\frac{1}{4}+1*\frac{1}{2}+2*\frac{1}{4}=1.$$

Визначимо дисперсії X і Y:

$$D_x=\Sigma(x_i-m_x)^2p_i=\frac{5}{18};$$

$$D_y = \Sigma(y_i - m_y)^2 p_i = \frac{1}{2}\ .$$

Визначимо кореляційний момент системи випадкових величин X, Y:

$$K_{xy} = \sum_{i=1}^{3}\sum_{j=1}^{3}(x_i - m_x)(y_j - m_y)p_{ij} = (0-\frac{1}{3})\left[(0-1)*\frac{1}{4}+(1-1)*\frac{1}{3}+(2-1)*\frac{1}{9}\right]+$$

$$+(1-\frac{1}{3})\left[0+(1-1)*\frac{1}{6}+(2-1)*\frac{1}{9}\right]+$$

$$+(2-\frac{1}{3})\left[(0-1)*0+(1-1)*0+(2-1)*\frac{1}{36}\right]=\frac{1}{6}\,.$$

Визначимо коефіцієнт кореляції:

$$r_{xy} = \frac{K_{xy}}{\sqrt{D_x D_y}} = \frac{\frac{1}{6}}{\sqrt{\frac{5}{18}*\frac{1}{6}}} = \frac{1}{\sqrt{5}}\,.$$

Функції випадкових величин

Функції однієї або декількох випадкових величин доводиться розглядати, коли аргументом деякої функції Y є система випадкових величин (X_1, X_2,…, X_n), закон розподілу яких відомий. Функція Y є випадковою величиною, закон розподілу якої необхідно визначити. У більшості задач для визначення числових характеристик функції декількох випадкових величин досить знати тільки числові характеристики аргументів.

1. Математичне сподівання суми двох залежних або незалежних випадкових величин X і Y дорівнює сумі їхніх математичних сподівань:

$$M[X+Y] = \Sigma(x_i+y_i)p_i = \Sigma x_i p_i + \Sigma y_i p_i = M[X] + M[Y]. \qquad (6.10)$$

Методом математичної індукції (узагальнення) для n доданків дістанемо:

$$M[\Sigma X_i] = \Sigma M[X_i],$$

математичне сподівання суми n випадкових величин дорівнює сумі їхніх математичних сподівань.

Математичне сподівання лінійної функції декількох випадкових величин

$$Y = \sum_{i=1}^{n} a_i X_i + b$$

дорівнює тій же лінійній функції від їхніх математичних сподівань:

$$M[Y] = M\left[\sum_{i=1}^{n} a_i X_i + b\right] = \sum_{i=1}^{n} a_i M[X_i] + b\,,$$

де a_i і b_i – невипадкові коефіцієнти.

2. Математичне сподівання добутку двох випадкових величин X і Y дорівнює добутку їхніх математичних сподівань плюс кореляційний момент. Запишемо вираз для кореляційного моменту:

$$K_{xy} = M[\overset{o}{Y}\overset{o}{X}] = M[(X-m_x)(Y-m_y)] = M[XY] - M[Xm_y] - M[Ym_x] + M[m_x m_y] =$$

$$= M[XY] - m_x m_y - m_y m_x + m_x m_y = M[XY] - m_x m_y\,,$$

звідки маємо

$$M[XY] = M[X] * M[Y] + K_{xy}. \quad (6.11)$$

Звідси кореляційний момент може бути виражений формулою

$$K_{xy}. = M[XY] - M[X] * M[Y]. \quad (6.12)$$

Якщо випадкові величини X і Y некорельовані, математичне сподівання добутку дорівнює добутку їхніх математичних сподівань:

$$M[XY] = M[X] * M[Y],$$

або для n незалежних співмножників:

$$M\left[\prod_{i=1}^{n} X_i\right] = \prod_{i=1}^{n} M[X_i].$$

3. Визначимо дисперсію суми двох випадкових величин X і Y:

$$D[X+Y] = M[((X+Y)-M(X+Y))^2] =$$
$$=M[X^2+2XY+Y^2 -2(X+Y)M(X+Y)+M^2(X+Y)] =$$
$$= M[X^2+2XY+Y^2]-2M(X+Y)M(X+Y)+M^2(X+Y) =$$
$$= M[X^2+2XY+Y^2]-(M[X]+ M[Y])^2 =$$
$$=M[X^2]+ M[Y^2]-M^2[X]- 2M[X]*M[Y]- M^2[Y] =$$
$$=D[X] + D[Y]+2M[XY] – 2M[X]M[Y] = D[X] + D[Y]+2K_{xy}$$

Таким чином, дисперсія суми двох випадкових величин X і Y дорівнює сумі їхніх дисперсій плюс подвоєний кореляційний момент

$$D[X+Y] = D[X] + D[Y] + 2K_{xy}, \quad (6.13)$$

Дисперсія суми декількох випадкових величин виражається формулою

$$D\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} D[X_i] + 2\sum_{i<j} K_{x_i x_j},$$

де $K_{x_i x_j}$ - кореляційний момент випадкових величин X_i і X_j.

Якщо X_i - некорельовані випадкові величини, то дисперсія їхньої суми дорівнює сумі їхніх дисперсій, тоді для суми n незалежних випадкових величин дисперсія визначиться в такий спосіб:

$$D[\Sigma X_i] = \Sigma D[X_i],$$

звідки середнє квадратичне відхилення суми $\sigma_\Sigma = \sqrt{\Sigma \sigma_i^2}$.

Дисперсія лінійної функції декількох випадкових величин

$$Y = \sum_{i=1}^{n} a_i X_i + b$$

виражається формулою

$$D[Y] = D\left[\sum_{i=1}^{n} a_i X_i + b\right] = \sum_{i=1}^{n} a_i^2 D[X_i] + 2\sum_{i<j} a_i a_j K_{x_i x_j}. \quad (6.14)$$

Якщо X_i - некорельовані випадкові величини, то

$$D[Y] = D\left[\sum_{i=1}^{n} a_i X_i + b\right] = \sum_{i=1}^{n} a_i^2 D[X_i].$$

4. Дисперсію добутку двох незалежних випадкових величин X і Y знаходимо за формулою

$$D[XY] = D[X]*D[Y] + M[X]^2 D[Y] + M[Y]^2 D[X]. \quad (6.15)$$

Приклад 6.4. На вимірювальний прилад надходить випадковий вектор (X, Y) з наступними характеристиками: m_x=-1; m_y=1; σ_x=2; σ_y=3;

r_{xy}=0,5. На виході приладу вимірюється величина $Z=(X-Y)^2$. Визначити математичне сподівання випадкової величини Z.

Розв'язання. Оскільки коефіцієнт кореляції не дорівнює нулю, випадкові величини X і Y корельовані. Перетворимо вираз Z:

$$M[Z] = M[(X-Y)^2] = M[X^2-2XY+Y^2]=$$

математичне сподівання суми випадкових величин дорівнює сумі математичних сподівань:

$$= M[X^2]- M[2XY]+ M[Y^2]=$$

другий початковий момент запишемо як суму дисперсії і квадрата математичного сподівання і врахуємо, що математичне сподівання добутку корельованих випадкових величин дорівнює добутку їхніх математичних сподівань плюс кореляційний момент:

$$= D_x+M^2[X] – 2(M[X]*M[Y]+K_{xy})+ D_y+M^2[Y] =$$
$$= 2^2+(-1)^2 – 2*(-1*1+2*3*0,5) + 3^2+1^2 = 11.$$

Приклад 6.5. Визначити математичне сподівання і дисперсію випадкової величини Z, якщо $Z=3+4(X-Y)$. Числові характеристики: m_x=-2; m_y=4; D_x=4; D_y=9; r_{xy}=-0,5.

Розв'язання. Оскільки коефіцієнт кореляції не дорівнює нулю, випадкові величини X і Y корельовані. Визначимо математичне сподівання Z:

$$M[Z] = M[3+4(X-Y)] = M[3]+4M[X+Y]=$$
$$= 3+4m_x+4m_y=3+4*(-2)+4*4=11.$$

Знайдемо дисперсію Z:

$$D[Z]= D[3+4(X-Y)] = D[3]+16D[X+Y]=$$
$$= 0+16(D[X]+D[Y]+2K_{xy}) =$$
$$=16*[4+9+2*(-0,5)(4*9] = 112,$$

де $K_{xy}=r_{xy}*\sqrt{D_x D_y}$.

Середнє квадратичне відхилення $\sigma_z = \sqrt{D_z} = \sqrt{112} = 10,6$.

Приклад 6.6. Випадкові величини X і Y незалежні і мають наступні характеристики: m_x=1; m_y=2; σ_x=1; σ_y=2. Обчислити математичне сподівання випадкових величин:

а) $U=X^2+2Y^2-XY-4X+Y+4$;

б) $V=(X+Y-1)^2$.

Розв'язання. а) Визначимо математичне сподівання випадкової величини U:

$$M[U]=M[X^2]+2M[Y^2]-M[XY]-4M[X]+M[Y]+M[4] =$$
$$= D[X]+M^2[X]+2(D[Y]+M_2[Y])-M[X]*M[Y]-4M[X]+M[Y]+4=$$
$$=1+1+2*(4+4)-1*2-4*1+2+4 = 18;$$

б) Визначимо математичне сподівання випадкової величини V:

$$M[V] = M[(X+Y-1)^2]= M[X^2+2Y-2+Y^2-2Y+1] =$$
$$= M[X^2]+2M[Y]-M[2]+M[Y^2]-2M[Y]+M[1] =$$
$$= D[X]+M^2[X]+2M[Y]-2+D[Y]+M^2[Y]-2M[Y]+1=$$
$$=1+1+4-2+4+4-4+1 = 9.$$

Приклад 6.7. Є випадкова величина X з математичним сподіванням m_x і дисперсією D_x. Знайти математичне сподівання і дисперсію наступних випадкових величин:

а) $Y=-X$;

б) $Z=X+2Y-1$;

в) $U=3X-Y+2Z-3$.

Розв'язання. а) Визначимо математичне сподівання випадкової величини Y:

$$а)\ M[Y]=M[-X]= -m_x;$$

знайдемо дисперсію випадкової величини Y

$$D[Y] = D[-X] = (-1)^2D_x = D_x;$$

б) Визначимо математичне сподівання випадкової величини Z:

$$M[Z]=M[X+2Y-1] = m_x -2m_x -1 = - m_x -1;$$

знайдемо дисперсію випадкової величини Z з урахуванням того, що математичне сподівання добутку корельованих випадкових величин дорівнює добутку їхніх математичних сподівань плюс кореляційний момент:

$$D[Z]=D[X+2Y-1] = D_x + 4D_y + 2Kxy =$$

кореляційний момент запишемо як різницю другого початкового моменту і добутку математичних сподівань і дістанемо

$$K_{xy} = M[X2Y] - M[X] * M[2Y]= M[X(-2X)] - M[X] * M[2Y] =$$
$$= -2M[X^2] + 2M^2[X] = -2(M[X^2] - M^2[X]) = -2D_x.$$

тоді дисперсія Z

$$D[Z] = D_x + 4D_x – 4D[X] = D[X].$$

в) Визначимо математичне сподівання випадкової величини Z:

$$M[U]= M[3X-Y+2Z-3] = 3m_x – m_y +2m_z -3 =$$
$$=3m_x + m_x +2(- m_x -1) -3 = 2m_x -5;$$

знайдемо дисперсію випадкової величини U з урахуванням формули (6.14)

$$D[U]= D[3X-Y+2Z-3] = D[3X] + D[-Y] + D[2Z] + 2(K_{xy}+K_{yz} + K_{xz});$$

визначимо кореляційні моменти

$$K_{xy} = M[3X(-Y)] - M[3X] * M[-Y]= 3M[X^2] - 3M[X] * M[X]=$$
$$= 3M[X^2] - 3M^2[X] = 3D_x;$$
$$K_{xz} = M[3X2Z] - M[3X] * M[2Z]= M[3X2(X+2Y-1)] - M[3X] * M[2X+4Y-2]=$$
$$= M[6X^2 -12X^2-6X)] - 3M[X] *(2M[X]+4M[-X]-2) =$$
$$=M[6X^2] - 12M[X^2]- 6M[X] - 6M^2[X] +12M^2[X] + 6 M[X] =$$
$$=-6 M[X^2] + 6 M^2[X] = -6 (M[X^2] - M^2[X]) = -6D_x.$$
$$K_{yz} = M[-Y2Z] - M[-Y] * M[2Z]= M[2X (X+2Y-1)] - M[X] * M[2(X+2Y-1)]=$$
$$= M[2X^2 -4X^2-2X] - M[X] * (2M[X]-4M[X]-2)=$$
$$=2M[X^2] -4M[X^2]-2M[X] - 2M^2[X] +4M^2[X]+2M[X]=$$
$$=-2M[X^2] + 2M^2[X]=-2(M[X^2] - M^2[X]) = -2D_x.$$

Визначимо тепер дисперсію випадкової величини U:

$$D[U]= D[3X-Y+2Z-3] = 9D[X] + D[X] + 4D[X] + 2(3D_x-6D_x-2D_x)=$$
$$=9D_x + D_x + 4D_x + 6D_x-12D_x-4D_x= 4 D_x.$$

Запитання для самоперевірки:

1. Що являє собою багатомірна випадкова величина?

2. Що являє собою функція розподілу системи двох випадкових величин? Перелічіть її властивості.

3. Перелічіть числові характеристики системи двох випадкових величин.

4. Що характеризує кореляційний момент системи двох випадкових величин?

5. Для чого використовується коефіцієнт кореляції?

6. Перелічіть теореми про числові характеристики.

7. Чому дорівнює середнє квадратичне відхилення добутку невипадкової величини C на випадкову величину X ?

8. Сформулюйте теорему додавання математичних сподівань для випадкових величин: а) залежних і незалежних; б) корельованих і некорельованих.

9. Чому дорівнює математичне сподівання добутку двох незалежних випадкових величин?

Задачі для самостійного розв'язання

6.1. Два стрілки незалежно один від іншого проводять по одному пострілу, кожний по своїй мішені. Випадкова величина X – число влучень першого стрільця, Y– число влучень другого стрільця. Імовірність влучення в мішень для першого стрільця дорівнює p_1=0,9, а для другого p_2=0,8. Побудувати функцію розподілу системи випадкових величин X і Y.

6.2. Для умов попереднього прикладу визначити числові характеристики випадкового вектора (X, Y).

6.3. Незалежні випадкові величини X і Y розподілені за нормальними законами з параметрами $m_x=2$; $m_y=-3$; $\sigma_x=1$; $\sigma_y=2$. Визначити імовірність події $A=\{X< m_x$ і $Y<m_y\}$.

6.4. Відомі математичне сподівання і дисперсія випадкової величини X: $m_x=2$; $D_x=3$. Знайти математичне сподівання і дисперсію випадкової величини $Y=3X-2$.

6.5. Випадкові величини X і Y мають математичні сподівання $m_x=-1$, $m_y=1$ і дисперсії $D_x=4$ і $D_y=9$. Знайти математичне сподівання випадкової величини $Z=3XY + 5$.

Тема 7. ЗАКОН ВЕЛИКИХ ЧИСЕЛ

Принцип практичної впевненості. Формулювання закону великих чисел

Якщо подія має дуже малу імовірність (відмінну від нуля), то в одиничному випробуванні ця подія може наступити і не наступити. Але так міркуємо ми тільки теоретично, а на практиці вважаємо, що подія, яка має малу імовірність, не наступить, і тому зневажаємо нею. Але у рамках математичної теорії не можна відповістити на запитання, якою повинна бути верхня межа імовірності, щоб можна було назвати «практично неможливими» події, імовірності яких не будуть перевищувати знайденої верхньої межі. Нехай, наприклад, робітник виготовляє на верстаті 100 виробів, з яких один в середньому виявляється бракованим. Очевидно, що імовірність браку дорівнює 0,01, але нею можна знехтувати і вважати робітника непоганим фахівцем. Але якщо будівельники будуть зводити будинки так, що з 100 будинків (у середньому) в одному будинку матиме місце руйнування даху, то навряд чи можна зневажити імовірністю такої події. Таким чином, у кожному окремому випадку слід виходити з того, наскільки важливі наслідки в результаті настання події.

Імовірність, якою можна знехтувати в даному дослідженні, називається **рівнем значущості**.

Сформулюємо принцип практичної впевненості: «Якщо випадкова подія має малу імовірність (наприклад, $p < 0{,}01$), то при одиничному випробуванні можна практично вважати, що ця подія не відбудеться, а якщо подія має імовірність, близьку до одиниці ($p > 0{,}99$), то практично при одиничному випробуванні можна вважати, що ця подія відбудеться напевно».

Основною закономірністю масових випадкових явищ є властивість стійкості середніх результатів. У широкому значенні слова під «законом великих чисел» розуміють відому з глибокої давнини властивість стійкості масових випадкових явищ, яка полягає в тому, що середній результат великої кількості випадкових явищ практично перестає бути випадковим і може бути передбаченим з достатньою визначеністю.

У вузькому значенні слова під «законом великих чисел» розуміють сукупність теорем, у яких встановлюється факт наближення середніх характеристик явища до деяких постійних величин у результаті великої кількості спостережень. Розглянемо деякі з них.

1. Лема Маркова. Якщо випадкова величина X не приймає від'ємних значень, то для будь-якого додатного числа справедлива нерівність

$$P\{X > \alpha\} \le \frac{M[X]}{\alpha}. \tag{7.1}$$

Доведення.

1) Нехай X — дискретна випадкова величина, задана рядом розподілу, причому $0 \le x_1 < x_2 < \ldots < x_n$.

x_i	x_1	x_2	…	x_n
p_i	p_1	p_2	…	p_n

Всі значення випадкової величини розіб'ємо на дві групи. До першої групи віднесемо значення, менші α (нехай це будуть $x_1, x_2,..., x_k$), до другої групи віднесемо всі інші значення, більші або рівні α ($x_{k+1}, x_{k+2},..., x_n$).
Як відомо, математичне сподівання дискретної випадкової величини визначається формулою

$$M[X]=x_1p_1 + x_2p_2 +... + x_kp_k + x_{k+l}p_{k+l} +... + x_np_n.$$

Відкинемо у правій частині формули перші k доданків. Оскільки $p_i > 0$, $x_i \geq 0$ і, крім того, при $x_i \geq \alpha$, $i \geq k+l$, матиме місце наступна нерівність: $M[X] \geq x_{k+l}p_{k+l} +... + x_np_n. \geq \alpha(p_{k+l} +... + p_n)$.

З того, що

$$p_{k+l} +... + p_n = P\{X=x_{k+1}\}+\ldots+P\{X=x_n\}=P\{X\geq\alpha\},$$

випливає, що:

$$M[X] \geq \alpha\, P\{X\geq\alpha\}.$$

Розділимо обидві частини останньої нерівності на α і дістанемо нерівність (7.1).

2) Нехай X - безперервна випадкова величина. Оскільки за умовою X не приймає від’ємних значень, її щільність імовірності $f(x) = 0$ при всіх $x < 0$. Тому

$$M[X] = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^{\infty} xf(x)dx \geq \int_{\alpha}^{\infty} xf(x)dx \geq \alpha\int_0^{\infty} xf(x)dx = \alpha P\{X \geq \alpha\},$$

розділимо на α, дістанемо нерівність (7.1). Лему доведено.

Приклад 7.1. Випадкова величина X задана рядом розподілу.

x_i	2	4	6	8	10	12
p_i	0,1	0,2	0,25	0,15	0,15	0,15

Визначити імовірність того, що випадкова величина X прийме значення, менше 11, використовуючи для цього нерівність Маркова.

Розв'язання. Виходячи з умови, запишемо:

$$P\{X<11\} = P\{X=2\} + P\{X=4\} + P\{X=6\}+ P\{X=8\}+ P\{X=10\}=$$
$$=0,1+0,2+0,25+0,15+0,15=0,85.$$

Використовуючи нерівність Маркова, дістанемо:

$$P\{X < 11\} \geq 1 - \frac{M[X]}{11} =$$
$$1 - \frac{2*0,1 + 4*0,2 + 6*0,25 + 8*0,15 + 10*0,15}{11} = 1 - 0,636 = 0,364$$
$$P\{X < 11\} \geq 0,364\,.$$

Приклад 7.2. Сума всіх внесків у банку становить 20000000 грн., а імовірність того, що випадково взятий внесок менше 100000 грн., дорівнює 0,8. Що можна сказати про число вкладників даного банку?

Розв'язання. Величину випадково взятого внеску позначимо X, а число всіх вкладників - n. Тоді з умови задачі виходить, що

$M[X] = \frac{20000000}{n}$, а імовірність, що $P\{X<100000\} = 0,8$.

за нерівністю Маркова $\quad P\{X<100000\} \geq 1-\frac{M[X]}{100000}$,

звідки маємо:

$$0{,}8 \geq 1-\frac{200000000}{n*100000}; \qquad \frac{200000000}{n*100000} \geq 0{,}2; \ 200 \geq n*0{,}2; \ n \leq 1000.$$

2. Нерівність Чебишева. Імовірність того, що відхилення випадкової величини X від її математичного сподівання за абсолютною величиною буде менше деякого додатного числа ε, обмежена знизу величиною

$$1-\frac{D[X]}{\varepsilon^2} \quad \text{або} \quad P\{|X-M[X]|<\varepsilon\} \geq 1-\frac{D[X]}{\varepsilon^2}. \tag{7.2}$$

Доведення.

Нехай маємо випадкову величину $(X\text{-}M[X])^2$. Оскільки вона приймає тільки додатні значення, до неї можна ужити лему Маркова, покладаючи в ній $\alpha=\varepsilon^2$, дістанемо:

$$P\{(X-M[X])^2<\varepsilon^2\} \geq 1-\frac{M[(X-M[X])^2]}{\varepsilon^2}.$$

Зазначимо, що $M[(X-M[X])^2]=D[X]$, а також врахуємо, що імовірність $P\{(X-M[X])^2<\varepsilon^2\}$ дорівнює імовірності $P\{|X-M[X]|<\varepsilon\}$, на цій підставі можемо записати:

$$P\{|X-M[X]|<\varepsilon\} \geq 1-\frac{D[X]}{\varepsilon^2}.$$

3. Теорема Чебишева. При необмеженому збільшенні числа n незалежних випробувань середня арифметична спостережуваних значень випадкової величини збігається за імовірністю до її математичного сподівання, тобто для деякого додатного ε:

$$P\{|\overline{X}-M[X]|<\varepsilon\}=1. \tag{7.3}$$

Ця теорема встановлює зв'язок між середньою арифметичною $\overline{X}$ спостережуваних значень випадкової величини X і її математичним сподіванням M[X].

Доведення.

Спостережувані значення X ($x_1, x_2, \ldots, x_n$) в чинність незалежності дослідів можна розглядати як незалежні випадкові величини, які мають однаковий розподіл (такий же, як в X) з однаковими параметрами – математичним сподіванням a = M[X] і дисперсією D[X]. За властивостями математичного сподівання і дисперсії можна записати:

$$M[\overline{X}]=M\left[\frac{X_1+X_2+\ldots+X_n}{n}\right]=\frac{1}{n}M[X_1+X_2+\ldots+X_n]=\frac{1}{n}(M[X_1]+M[X_2]+\ldots+M[X_n])=$$

$$=\frac{1}{n}(a+a+\ldots+a)=\frac{1}{n}na=a.$$

$$D[\overline{X}]=D\left[\frac{X_1+X_2+...+X_n}{n}\right]=\frac{1}{n^2}D[X_1+X_2+...+X_n]=\frac{1}{n^2}(D[X_1]+D[X_2]+...+D[X_n])=$$
$$=\frac{1}{n^2}nD[X]=\frac{D[X]}{n}.$$

Скористаємося нерівністю Чебишева для випадкової величини X і підставимо в нього отримані параметри:

$$P\{\left|\overline{X}-a\right|<\varepsilon\}\geq 1-\frac{D[X]}{n\varepsilon^2}\ . \tag{7.4}$$

Якщо тепер в отриманій нерівності взяти як завгодно мале додатне ε і необмежено збільшити n, одержимо

$$\lim_{n\to\infty}\left(1-\frac{D[X]}{n\varepsilon^2}\right)=1,$$

що й доводить теорему Чебишева.

З теореми Чебишева випливає важливий практичний висновок: невідоме значення математичного сподівання випадкової величини ми вправі замінити середнім арифметичним значенням, отриманим на підставі досить великої кількості дослідів. При цьому чим більше проведено дослідів, тим з більшою імовірністю можна чекати, що пов'язана з цією заміною помилка ($\overline{X}-a$) не перевершить задану величину ε. При відомому значенні дисперсії D[X], наприклад, за заданим значенням імовірності $P\{\left|\overline{X}-a\right|<\varepsilon\}$ і максимальній припустимій помилці ε можна визначити число необхідних дослідів n; або за заданою імовірністю P і числом дослідів n визначити величину помилки ε; а також, за заданим ε і n можна визначити межу імовірності події $\left|\overline{X}-a\right|<\varepsilon$.

Приклад 7.3. Дисперсія випадкової величини X дорівнює 4. Скільки потрібно зробити незалежних дослідів, щоб з імовірністю не менше 0,9 можна було очікувати, що середнє арифметичне значення цієї випадкової величини буде відрізнятися від її математичного сподівання менше ніж на 0,5?

Розв'язання. За умовою задачі ε = 0,5; $P\{\left|\overline{X}-a\right|<0{,}5\}\geq 0{,}9$; визначити n. Скористаємося формулою (7.4), врахуємо при цьому, що $1-\frac{D[X]}{n\varepsilon^2}=0{,}9$:

$$P\{\left|\overline{X}-a\right|<\varepsilon\}\geq 1-\frac{D[X]}{n\varepsilon^2}$$
$$n=\frac{D[X]}{0{,}1\varepsilon^2}=\frac{4}{0{,}1*0{,}5^2}=160.$$

Коли використати твердження, що середнє арифметичне розподілено приблизно нормально, то дістанемо:

$$P\{\left|\overline{X}-a\right|<\varepsilon\}=2\Phi_0\varepsilon\sqrt{\frac{n}{\sigma}}\geq 0{,}9,\ \text{звідки}\ \varepsilon\sqrt{\frac{n}{\sigma}}\geq 1{,}645,\ n\geq 49.$$

Приклад 7.4. Дисперсія випадкової величини X дорівнює 5. Зроблено 100 незалежних дослідів, за якими обчислено середнє значення $\overline{X}$. Замість невідомого значення математичного сподівання *a* прийнято середнє значення $\overline{X}$. Визначити максимальну величину помилки, що допускається при цьому, з імовірністю не менше 0,8.

Розв'язання. За умовою n = 100; $P\{\left|\overline{X} - a\right| < \varepsilon\} \geq 0{,}8$; визначити ε. Використаємо формулу (7.4):

$$P\{\left|\overline{X} - a\right| < \varepsilon\} \geq 1 - \frac{D[X]}{n\varepsilon^2},$$

із співвідношення $1 - \frac{D[X]}{n\varepsilon^2} = 0{,}8$ визначимо ε:

$$\varepsilon^2 = \frac{D[X]}{0{,}2n} = \frac{5}{0{,}2 * 100} = 0{,}25; \quad \varepsilon = 0{,}5.$$

4. Теорема Бернуллі. При необмеженому зростанні числа незалежних випробувань n відносна частота $\frac{m}{n}$ появи події A збігається за імовірністю до її імовірності P, тобто

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n} - p\right| \leq \varepsilon\right) = 1. \qquad (7.5)$$

За допомогою цієї теореми встановлюється зв'язок між відносною частотою (частостью) події і її імовірністю. Вона була доведена Я.Бернуллі (опублікована в 1713 р.) і поклала початок теорії імовірностей як науці.

Доведення.

Відносна частота $\frac{m}{n}$ є випадковою величиною, а отже характеризується математичним сподіванням і дисперсією:

$$M\left[\frac{m}{n}\right] = p; \quad D\left[\frac{m}{n}\right] = \frac{pq}{n}.$$

Запишемо нерівність Чебишева для випадкової величини $\frac{m}{n}$:

$$P\left\{\left|\frac{m}{n} - p\right| < \varepsilon\right\} \geq 1 - \frac{D\left[\frac{m}{n}\right]}{\varepsilon^2}.$$

Остаточно маємо:

$$P\left\{\left|\frac{m}{n} - p\right| < \varepsilon\right\} \geq 1 - \frac{pq}{n\varepsilon^2}. \qquad (7.6)$$

Яким би малим не було число ε, при n→∞ величина дробу $\frac{pq}{n\varepsilon^2} \to 0$, а $P\left\{\left|\frac{m}{n} - p\right| < \varepsilon\right\} \to 1$.

З теореми Бернуллі випливає, що при досить великій кількості випробувань відносна частота $\frac{m}{n}$ появи події практично втрачає свій випадковий характер, наближаючись до постійної величини P — імовірності даної події. У цьому й полягає принцип практичної впевненості.

Незважаючи на те, що при необмеженому зростанні числа незалежних випробувань різниця $\left|\frac{m}{n}-p\right|$ може виявитися як завгодно малою, все таки не можна сказати, що $\lim\limits_{n\to\infty}\frac{m}{n}=P$. Таке твердження було б невірним, оскільки тут не виконуються необхідні умови, які входять до складу визначення поняття границі. Справді, може статися, що подія A буде відбуватися при всіх наступних випробуваннях, починаючи з деякого номера n > N і тоді $\lim\limits_{n\to\infty}\frac{m}{n}=P$, але не виключений і випадок, коли починаючи з деякого номера n > N, подія A не буде відбуватися при жодному випробуванні, тоді $\lim\limits_{n\to\infty}\frac{m}{n}=0$.

Таким чином, при необмеженому числі незалежних випробувань може статися, що $\frac{m}{n}$ p, але цього може і не статися. Тоді виникає запитання про те, яка ж імовірність того, що $\frac{m}{n}$ p? З теореми Бернуллі відповіді на це запитання не випливає, але в більш глибоких дослідженнях з теорії імовірностей доводиться, що при $n\to\infty$ $P\left\{\frac{m}{n}\to p\right\}=1$. Отже $\frac{m}{n}$ p не за типом границі, а за імовірністю.

Приклад 7.5. З метою встановлення частки браку продукції було перевірено за схемою зворотної вибірки 1000 одиниць. Яка імовірність того, що встановлена цією вибіркою частка браку за абсолютною величиною буде відрізнятися від частки браку в усій партії не більше ніж на 0,01, коли відомо, що в середньому на кожні 10000 виробів приходиться 500 бракованих?

Розв'язання. За умовою задачі число незалежних випробувань n = 1000, частота $p=\frac{500}{10000}=0{,}05$, q=1-p=0,95, ε=0,01. Треба визначити імовірність події $P\left\{\left|\frac{m}{n}-p\right|<0{,}01\right\}$. Скористаємося формулою (7.6), дістанемо:

$$P\left\{\left|\frac{m}{n}-p\right|<0{,}01\right\}\ge 1-\frac{pq}{n\varepsilon^2}=1-\frac{0{,}05*0{,}95}{1000*0{,}0001}=0{,}527.$$

Таким чином, з імовірністю не менше 0,527 можна чекати, що вибіркова частка браку (відносна частота появи браку) буде відрізнятися від частки браку в усій продукції (від імовірності браку) не більше ніж на 0,01.

Приклад 7.6. При штампуванні деталей імовірність браку становить 0,05. Скільки потрібно перевірити деталей, щоб з імовірністю не менше 0,95 можна було очікувати, що відносна частота бракованих виробів буде відрізнятися від імовірності браку менше ніж на 0,01?

Розв'язання. З умови задачі відомо, що p = 0,05; q = 0,95; ε= 0,01, $P\left\{\left|\frac{m}{n}-p\right|<0{,}01\right\}\geq 0{,}95$. Треба визначити n. Запишемо рівність

$$1-\frac{pq}{n\varepsilon^2}=0{,}95,$$

звідки дістанемо $n=\frac{pq}{0{,}05\varepsilon^2}=\frac{0{,}05*0{,}95}{0{,}05*0{,}0001}=9500.$

Запитання для самоперевірки:

1. Що називається законом великих чисел? Поясніть смисл цієї назви.
2. Яка роль закону великих чисел у теорії імовірностей?
3. У чому полягає принцип практичної впевненості?
4. Поясніть смисл поняття «рівень значущості».
5. Сформулюйте теорему Чебишева і поясніть, в чому полягає її практичний зміст.
6. Сформулюйте теорему Бернуллі і поясніть, в чому полягає її практичний зміст.
7. Чи можна стверджувати, що при нескінченно великій кількості дослідів n частота події p* дорівнює імовірності цієї події p? Обґрунтуйте відповідь.

Задачі для самостійного розв'язання

7.1. Вага виробу, що виготовляється підприємством, є випадковою величиною з математичним сподіванням 90 г і дисперсією 0,0225. Визначити імовірність того, що відхилення ваги виробу від її середнього значення за абсолютною величиною не перевищить 0,4 г. Для розв'язання використати нерівність Чебишева.

7.2. З 1000 виробів, що надходять у складальний цех, випадковим способом вибрали 200 виробів для контролю. Серед них виявилося 25 бракованих. Прийнявши частку бракованих виробів з контрольної партії як імовірність виготовлення бракованого виробу, оцінити імовірність того, що у всій партії виробів, які надійшли у складальний цех, бракованих виявиться не менше 10 % і не більше 15 %. Для розв'язання використати теорему Бернуллі.

Змістовий модуль 2. ЕЛЕМЕНТИ МАТЕМАТИЧНОЇ СТАТИСТИКИ

Тема 8. ОБРОБКА СТАТИСТИЧНИХ ДАНИХ

Математична статистика розробляє методи обробки результатів спостережень масових випадкових явищ з метою виявлення наявних у них закономірностей. Збір статистичних даних проводять за спеціальними правилами статистичного спостереження.

Суть **вибіркового методу** полягає в тому, що висновки, зроблені на основі вивчення частини сукупності (випадкової вибірки), можна поширити на всю сукупність (генеральну сукупність). Числові характеристики генеральної сукупності називаються генеральними параметрами. До них відносять математичне сподівання і дисперсію, що є параметрами розподілу досліджуваної ознаки. Їхні теоретичні значення ніколи не відомі, але їх можна оцінити за значеннями вибіркових характеристик.

Оцінкою параметра називається числова характеристика, отримана в результаті обробки випадкової вибірки.

Статистичними аналогами параметрів розподілу є генеральна середня $\overline{X}$ і генеральна дисперсія $\sigma^2_{ген}$. Розраховані за даними вибірки числові характеристики позначають $\tilde{X}$ або $\overline{X}_{виб}$, $\sigma_{виб}$ або $S_{виб}$.

При здійсненні вибірки можливі помилки спостереження, до яких відносять помилки реєстрації і помилки репрезентативності. **Помилки реєстрації** виникають через неточності й похибки при одержанні відомостей про одиниці сукупності, в результаті чого дійсне значення досліджуваної ознаки не збігається з її зареєстрованим значенням. **Помилки репрезентативності** являють собою різницю між вибірковими і генеральними характеристиками досліджуваної сукупності. Їх підрозділяють на систематичні й випадкові. **Систематичні помилки** пов'язані з тим, що структура вибірки відрізняється від структури генеральної сукупності. Звичайно це пов'язане з порушенням випадковості відбору. **Випадкові помилки** репрезентативності пояснюються тим, що досліджується тільки частина сукупності. Таким чином, наявність випадкової помилки споконвічно властива вибірковому методу.

Введемо ряд визначень.

Об'єктом спостереження називається сукупність предметів або явищ, об'єднаних якою-небудь спільною ознакою або властивістю якісного або кількісного характеру X. Всякий об'єкт статистичного спостереження складається з окремих елементів — **одиниць спостереження**. Для вивчення характерних властивостей об'єкта випадковим способом відбирають з усієї сукупності обмежене число одиниць спостереження.

Генеральною сукупністю називають об'єкт спостереження, що містить всю сукупність одиниць, з яких проводиться вибірка.

Вибірковою сукупністю, або просто **вибіркою**, називають сукупність одиниць спостереження, випадково відібраних з генеральної сукупності.

Розрізняють два типи випадкових вибірок: властиво випадкова повторна вибірка (схема повернутої кулі) і властиво випадкова безповторна вибірка (схема неповернутої кулі). Вибір схеми відбору залежить від характеру досліджуваного об'єкта.

Обсягом сукупності (вибіркової або генеральної) називають число одиниць цієї сукупності. Число одиниць вибіркової сукупності позначають n, а генеральної - N.

У результаті статистичного спостереження отримують набір значень спостережуваної ознаки - дані. Значення ознаки при переході від одного елемента до іншого варіюють (змінюються).

Варіантой називають кожне окреме значення досліджуваної ознаки

$$x_1, x_2, \ldots, x_n \ldots$$

Частотою називають число, що показує, скільки разів зустрічається в сукупності та або інша варіанта. Частоти позначають відповідно

$$m_1, m_2, \ldots, m_n \ldots$$

Первинною статистичною сукупністю називається невпорядкований набір значень досліджуваної ознаки, отриманих в результаті спостереження.

Першим кроком в обробці статистичних даних є впорядкування отриманих значень ознаки в порядку зростання або убування, тобто побудова варіаційного ряду.

Варіаційним рядом називається таблиця, в одному рядку якої розташовуються варіанти $x_1, x_2, \ldots, x_n$ у зростаючому або убутному порядку, а в другому — відповідні їм частоти $m_1, m_2, \ldots, m_n \ldots$

x_i	x_1	x_2	…	x_n
m_i	m_1	m_2	…	m_n

де $\sum_{i=1}^{n} m_i = n$.

Варіація досліджуваної ознаки може бути дискретною і безперервною.

Дискретною називається варіація, при якій окремі значення ознаки (варіанти) відрізняються одне від одного на деяку скінченну величину (звичайно ціле число).

Безперервною називається варіація, при якій значення ознаки можуть відрізнятися одне від іншого на як завгодно малу величину.

Замість абсолютних значень частот m_i звичайно використовують відносні – p^*_i. Для одержання відносних частот необхідно відповідну частоту розділити на суму всіх частот:

$$p_1^* = \frac{m_1}{\sum_{i=1}^{n} m_i}, \qquad p_2^* = \frac{m_2}{\sum_{i=1}^{n} m_i}, \quad \ldots, \qquad p_n^* = \frac{m_n}{\sum_{i=1}^{n} m_i} \tag{8.1}$$

Сума всіх відносних частот дорівнює одиниці:

$$\sum_{i=1}^{n} p_i^* = 1. \tag{8.2}$$

Якщо варіація ознаки має безперервний характер, то весь діапазон отриманих значень розділяють на інтервали так, що кінець одного інтервалу є початком наступного. При цьому частоти належать не до окремого значення ознаки, а до інтервалу значень. Звичайно як значення інтервалу приймають його середину.

У вибірці, що має дискретний характер варіації, число одиниць спостереження n може виявитися дуже великим. У результаті варіаційний ряд може виявитися громіздким і незручним для обробки. У такому випадку діапазон отриманих значень також розділяють на інтервали. Довжину і-го інтервалу позначають k_i і знаходять за формулою

$$k_i = x_{imax} - x_{imin}, \tag{8.3}$$

де x_{imax} і x_{imin} - відповідно верхня і нижня межі і -го інтервалу.

При угрупованні даних з метою побудови інтервального варіаційного ряду завжди встає питання про вибір оптимального числа інтервалів. Це пов'язано з тим, що занадто велике число інтервалів зробить варіаційний ряд громіздким, а занадто мале призведе до огрубіння результатів статистичного аналізу. Для визначення оптимальної довжини інтервалу рекомендується використовувати формулу Старджеса:

$$k = \frac{x_{\max} - x_{\min}}{1 + 3{,}322\lg n}, \tag{8.4}$$

де x_{max} і x_{min} - відповідно найбільше і найменше значення варіантів ряду; n – число одиниць сукупності.

Для наочності варіаційний ряд можна зобразити графічно у вигляді полігона розподілу, кумулятивної кривої, або гістограми.

Полігон розподілу (дослівно — многокутник розподілу) будують в прямокутній системі координат. Величину ознаки відкладають на осі абсцис, частоти або відносні частоти - на осі ординат (рис. 8.1).

Кумулятивна крива (кумулята) утворюється при зображенні варіаційного ряду з накопиченими відносними частотами в прямокутній системі координат. Накопичена частота певної варіанти утворюється підсумовуванням всіх частот варіант, які передують даній, з частотою цієї варіанти. При побудові кумуляти дискретної ознаки на осі абсцис відкладають значення ознаки (варіанти). Ординатами служать вертикальні відрізки, довжина яких пропорційна накопиченій відносній частоті тієї або іншої варіанти Σp_i^* (рис. 8.2). З'єднавши вершини ординат прямими лініями, отримаємо ламану (криву) кумуляту. При $\Delta x_i \to 0$ кумулята прагне до безперервної кривої, що являє собою функцію розподілу досліджуваної ознаки.

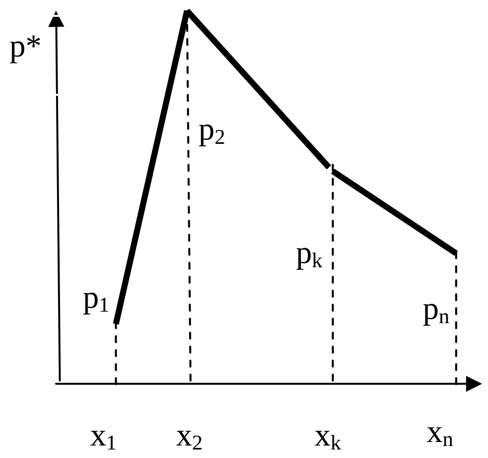


Рис. 8.1 - Полігон розподілу

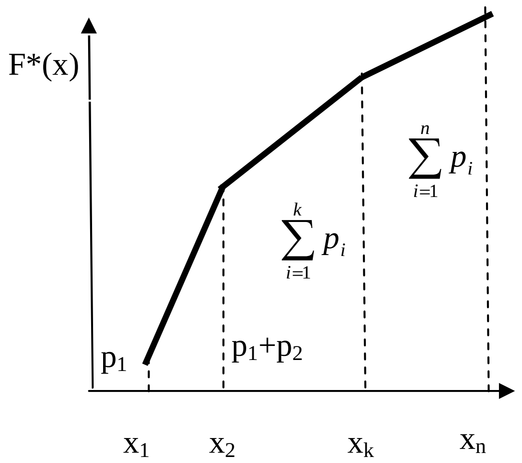


Рис. 8.2 - Кумулятивна крива

Гістограму розподілу будують аналогічно полігону в прямокутній системі координат. При побудові гістограми на осі абсцис вибирають відрізки, що відповідають інтервалам, на яких будують прямокутники з площею, пропорційною відносним частотам інтервалів p_i^*. З умови побудови гістограми випливає, що вся її площа дорівнює 1 (рис. 8.3). Очевидно також, що при $\Delta x_i \to 0$ гістограма прагне до безперервної залежності, що являє собою щільність розподілу випадкової величини.

Визначення закону розподілу спостережуваної ознаки за статистичним даними

Одним із завдань математичної статистики є отримання оцінок числових характеристик досліджуваної ознаки. Іншим завданням є визначення закону розподілу досліджуваної ознаки за статистичним даними. Оскільки ці дані завжди обмежені, то виникає задача згладжування або вирівнювання статистичних рядів за допомогою аналітичних виразів, що є найбільш компактним виразом закономірності. Статистичний (варіаційний) ряд дозволяє побудувати шукані статистичні закони розподілу. Зокрема, для того, щоб знайти статистичну функцію розподілу F*(x), досить підрахувати число варіант, в яких ознака X прийняла значення менші x, тобто X<x. Якщо число таких варіант m(x), а обсяг сукупності дорівнює n, то

$$F^*(x) = \frac{m(x)}{n}. \tag{8.4}$$

Якщо результати спостереження зведені в групований статистичний ряд, то на його підставі будують гістограму з урахуванням того, що частота появ ознаки на i-му інтервалі буде

$$p^* = \frac{m_i}{n},$$

де m_i – кількість появи x на i-му інтервалі.

Для оформлення статистичного ряду у вигляді гістограми на осі абсцис відкладають інтервали ряду, а потім на кожному інтервалі будують прямокутник, площа якого дорівнює p^*_i.

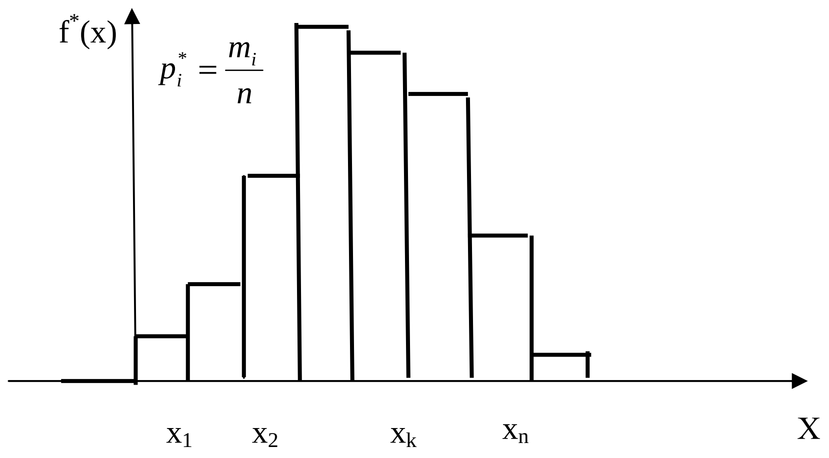


Рис. 8.3 - Гістограма

Завдання вирівнювання статистичних рядів зводиться до підбору теоретичної кривої розподілу, яка виражає лише істотні риси статистичного матеріалу. Якщо клас функцій, які описують розподіл, відомий, то задача зводиться до раціонального вибору параметрів розподілу. Наприклад, гістограма, наведена на рис. 8.3, наводить на думку про нормальний розподіл ознаки X:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-m)^2}{2*\sigma^2}\right\}.$$

Тоді задача зводиться до відшукання двох параметрів – математичного сподівання m і σ, які оцінюються за вибірковою середньою $\tilde{X}$, і вибірковим середнім квадратичним відхиленням $\sigma_{виб}$.

Одним з методів розв'язання поставленої задачі є метод моментів (початкових і центральних), що був запропонований в 1894 р. англійцем Пірсоном. Відповідно до цього методу числові параметри розподілу вибираються з таким розрахунком, щоб кілька найважливіших числових характеристик (моментів) дорівнювали їхнім статистичним оцінкам. Зокрема, математичне сподівання і дисперсію приймають рівними їхнім статистичним оцінкам - вибірковій середній і вибірковій дисперсії:

$$m_x = \tilde{X}, \qquad D_x = \sigma^2_{виб}.$$

Іншим методом визначення статистичних оцінок параметрів розподілу є метод максимальної правдоподібності, запропонований в 1912 р. англійцем Фішером.

Приклад 8.1. За результатами розпродажу товарів 26 продавцями побудований варіаційний ряд:

Число продажів (x_i)	9	12	13	14	15	16	17	19	21	23	27
Число продав-ців(m_i)	1	2	3	6	5	3	2	1	1	1	1

Знайти статистичну функцію розподілу F*(x).

Розв'язання. Очевидно, що шукана статистична функція розподілу для всіх x<9 дорівнює нулю (тому що число продажів не приймало значення менш дев'яти)

$$F^*(9)=0.$$

Визначимо частоту числа продажів x<12, до них відноситься число продажів 9, причому m(9)=1, тоді

$$F^*(12)=\frac{m(9)}{n}=\frac{1}{26}\approx 0{,}04\,.$$

Визначимо частоту числа продажів x<13, до них відноситься число продажів 9 і 12, причому m(9)=1, а m(12)=2, тоді

$$F^*(13)=\frac{m(9)}{n}+\frac{m(12)}{n}=\frac{1}{26}+\frac{2}{26}\approx 0{,}04+0{,}08=0{,}12\,.$$

Аналогічно отримаємо інші значення статистичної функції розподілу і представимо їх у табличному вигляді:

x_i	9	12	13	14	15	16	17	19	21	23	27	x>27
$F(x_i)$	0	0,04	0,12	0,23	0,46	0,65	0,77	0,85	0,88	0,92	0,96	1

Графік статистичної функції розподілу поданий на рис. 8.4.

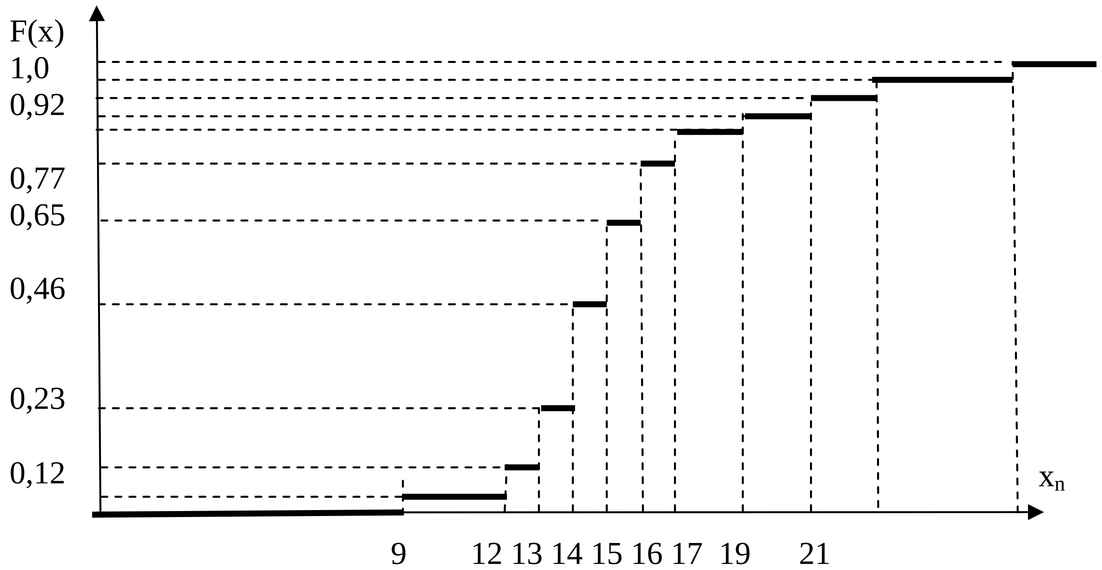


Рис. 8.4 - Статистична функція розподілу

Приклад 8.2. З метою дослідження точності приладу зроблено 500 вимірювань помилки. Результати вимірювань зведені в групований варіаційний ряд:

групи	-4 ...-3	-3 ...-2	-2 ...-1	-1 ...0	0 ...1	1 ...2	2 ...3	3 ...4
m_i	6	25	72	133	120	88	46	10

Побудувати гістограму розподілу і визначити параметри розподілу.

Розв'язання. Визначимо частоти для кожного розряду групованого варіаційного ряду, користуючись формулою

$$p_i^* = \frac{m_i}{n},$$

де m_i – число значень помилки X, що потрапили в i-у групу; n – число зроблених вимірів, n = 500.

групи	-4 …-3	-3 …-2	-2 …-1	-1 …0	0 …1	1 …2	2 …3	3 …4
m_i	6	25	72	133	120	88	46	10
p_i*	0,012	0,05	0,144	0,266	0,24	0,176	0,092	0,02

Для побудови гістограми знайдемо значення щільності частот для кожної групи значень за формулою

$$f_i^* = \frac{p_i^*}{l},$$

де l - довжина групи, l= 1.

групи	-4 …-3	-3 …-2	-2 …-1	-1 …0	0 …1	1 …2	2 …3	3 …4
m_i	6	25	72	133	120	88	46	10
p_i*	0,012	0,05	0,144	0,266	0,24	0,176	0,092	0,02
f_i*	0,012	0,05	0,144	0,266	0,24	0,176	0,092	0,02

Побудуємо графік гістограми (рис. 8.5).

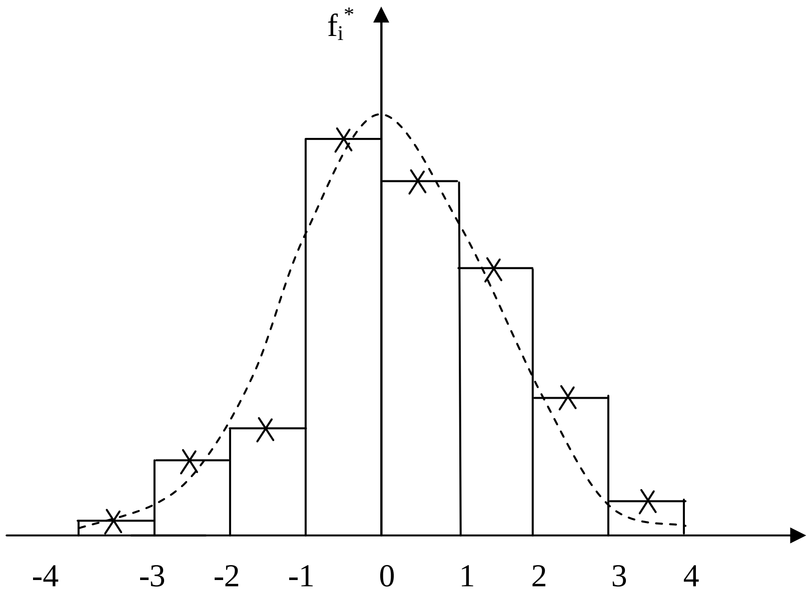


Рис. 8.5 - Гістограма

З вигляду гістограми можна припустити, що її можна згладити за допомогою нормального закону (припущення підтверджується тим, що досліджувана випадкова величина є помилкою вимірювання, а отже розподілена нормально), щільність розподілу якого визначається виразом

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}.$$

Параметри m і σ, що входять у вираз щільності f(x), підбирають так, щоб найкраще погодити аналітичний вираз зі статистичним розподілом. Визначимо вибіркове середнє $\tilde{X}$ і вибіркове середнє квадратичне відхилення $\sigma_{выб}$, користуючись даними групованого варіаційного ряду. Як значення x_i виберемо середину і-ї групи і цьому значенню поставимо у відповідність як імовірність його частоту p_i*, тоді отримаємо

$$\tilde{X} = \sum_{i=1}^{8} x_i p_i^* = -3{,}5*0{,}012 - 2{,}5*0{,}05 - 1{,}5*0{,}144 - 0{,}5*0{,}266 + \\ + 0{,}5*0{,}24 + 1{,}5*0{,}176 + 2{,}5*0{,}092 + 3{,}5*0{,}02 = 0{,}168;$$

$$\tilde{\alpha}_2 = \sum_{i=1}^{8} x_i^2 p_i^* = (-3{,}5)^2*0{,}012 + (-2{,}5)^2*0{,}05 + (-1{,}5)^2*0{,}144 + (-0{,}5)^2*0{,}266 + \\ + (0{,}5)^2*0{,}24 + (1{,}5)^2*0{,}176 + (2{,}5)^2*0{,}092 + (3{,}5)^2*0{,}02 = 2{,}126;$$

$$\sigma^2_{выб} = \tilde{\alpha}_2 - \tilde{X}^2 = 2{,}126 - (0{,}168)^2 = 2{,}098;$$

$$\sigma_{выб} = \sqrt{\sigma^2_{выб}} = \sqrt{2{,}098} = 1{,}448.$$

Маємо розподіл

$$f^*(x) = \frac{1}{1{,}448\sqrt{2\pi}} \exp\left\{-\frac{(x-0{,}168)^2}{2*2{,}126}\right\},$$

користуючись яким, підрахуємо значення f*(x) на межах груп:

$f_i(-4) = 0{,}0045$; $f_i(1) = 0{,}2343$;
$f_i(-3) = 0{,}0256$; $f_i(2) = 0{,}1244$;
$f_i(-2) = 0{,}0895$; $f_i(3) = 0{,}0435$;
$f_i(-1) = 0{,}1986$; $f_i(4) = 0{,}0087$;
$f_i(0) = 0{,}274$.

Відкладемо на графіку отримані точки і проведемо плавну криву (рис. 8.5).

Числові характеристики варіаційного ряду

Однією з найважливіших характеристик варіаційного ряду є його середнє значення. Розрізняють середню арифметичну зважену і незважену. Середня арифметична зважена визначається як відношення суми добутків варіантів на відповідні ним частоти до суми всіх частот:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i m_i}{\sum_{i=1}^{k} m_i} \quad \text{або} \quad \bar{x} = \sum_{i=1}^{k} x_i p_i^* \tag{8.5}$$

де m_i – частоти варіаційного ряду; p_i* - відносні частоти; k – число груп.

Для розрахунку незваженої середньої арифметичної використовуємо формулу

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{8.6}$$

де n - число елементів варіаційного ряду, $n=\Sigma m_i$.

Залежно від особливостей досліджуваного явища крім арифметичної використовують геометричну, гармонійну, квадратичну, кубічну та інші середні величини.

Розглянемо ряд характеристик, що використовуються для вимірювання варіації ознаки.

Варіаційний розмах R, або широта розподілу - різниця між найбільшим і найменшим значеннями варіаційного ряду:

$$R = x_{max} - x_{min} \tag{8.7}$$

Варіаційний розмах застосовують для приблизної оцінки варіації, тому що він говорить тільки про відстань між найбільшим і найменшим значеннями варіантів варіаційного ряду. Більш точну міру варіації визначають за допомогою середнього лінійного відхилення, дисперсії і стандартного відхилення (або середнього квадратичного відхилення).

Середнє лінійне відхилення, або просте середнє відхилення (позначається d) являє собою середнє арифметичне абсолютних значень відхилень варіант від середньої. Обчислюють середнє лінійне відхилення незважене або зважене:

$$d = \frac{\sum_{i=1}^{n} | x_i - \overline{X} |}{n}, \qquad d = \frac{\sum_{i=1}^{k} | x_i - \overline{X} | m_i}{\sum_{i=1}^{k} m_i}. \tag{8.8}$$

Дисперсія варіаційного ряду – це середня арифметична квадрата відхилення значень ознак ряду від їхньої середньої арифметичної. Дисперсія незважена і зважена обчислюється за формулами

$$\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \overline{X})^2}{n}, \qquad \sigma^2 = \frac{\sum_{i=1}^{k} (x_i - \overline{X})^2 m_i}{\sum_{i=1}^{k} m_i}. \tag{8.9}$$

Стандартне відхилення варіаційного ряду визначається як арифметичне значення квадратного кореня з дисперсії:

$$\sigma = \sqrt{\sigma^2} \;. \tag{8.10}$$

Дисперсія - це міра розсіювання варіантів щодо середньої арифметичної. Чим більше варіація, тим далі від середньої перебувають можливі значення ознак.

Коефіцієнт варіації V являє собою частку стандартного відхилення від середнього значення варіаційного ряду:

$$V = \frac{\sigma}{\tilde{X}} \quad . \tag{8.11}$$

Таким чином, коефіцієнт варіації є відносною мірою варіації, тоді як стандартне відхилення - абсолютна міра розсіювання варіантів ряду. Чим менше значення коефіцієнта варіації, тим однорідніше сукупність за досліджуваною ознакою.

Приклад 8.3. Обчислити числові характеристики для варіаційного ряду з прикладу 8.1:

Число продажів (x_i)	9	12	13	14	15	16	17	19	21	23	27
Число продавців(m_i)	1	2	3	6	5	3	2	1	1	1	1

Розв'язання. Визначимо зважену середню арифметичну за формулою (8.5)

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i m_i}{\sum_{i=1}^{k} m_i}$$

Скористаємося для цього таблицею Microsoft Excel. Розрахунки показані в рядках 1-5 таблиці на рис. 8.6. і 8.7

$$\bar{x} = \frac{9*1+12*2+13*3+14*6+15*5+16*3+17*2+19+21+23+27}{26} = 15{,}5$$

Визначимо варіаційний розмах ряду

$$R = x_{max} - x_{min} = 27\text{-}9 = 18.$$

Знайдемо середнє лінійне відхилення. Для обчислення абсолютних значень різностей $|x_i - \bar{x}|$ скористаємось функцією ABS(B1-E5), аргументом якої є дійсне число. Розрахунок показаний у рядках 7-10 таблиці на рис. 8.6 і 8.7.

$$d = \frac{\sum_{i=1}^{k} |x_i - \bar{x}| m_i}{\sum_{i=1}^{k} m_i} = \frac{|9-15{,}5| * + |12-15{,}5| * 2 + |13-15{,}5| * 3 + |14-15{,}5| * 6 + |15-15{,}5| * 5}{26} +$$

$$+ \frac{(16-15{,}5)*3+(17-15{,}5)*2+(19-15{,}5)+(21-15{,}5)+(23-15{,}5)+(27-15{,}5)}{26} = 2{,}5$$

Дисперсію обчислимо за формулою (8.9). Для розрахунку скористаємось функцією =СТЕПЕНЬ(В7;2). Аргументи цієї функції - основа степені (будь-яке дійсне число) і показник степені:

$$\sigma^2=\frac{\sum_{i=1}^{k}(x_i-\bar{x})^2 m_i}{\sum_{i=1}^{k} m_i}=\frac{(9-15{,}5)^2+(12-15{,}5)^2*2+(13-15{,}5)^2*3+(14-15{,}5)^2*6+(15-15{,}5)^2*5}{26}+$$

$$+\frac{(16-15{,}5)^2*3+(17-15{,}5)^2*2+(19-15{,}5)^2+(21-15{,}5)^2+(23-15{,}5)^2+(27-15{,}5)^2}{26}=12{,}94.$$

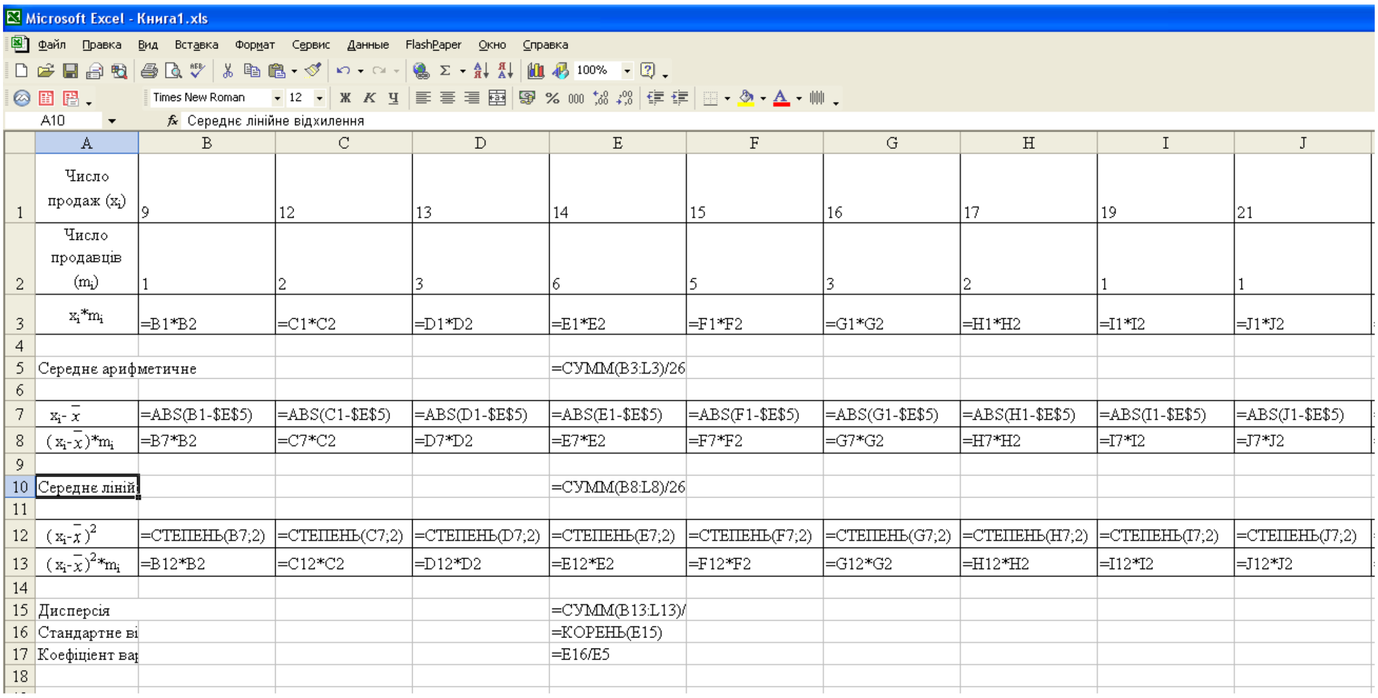
Розрахунки дисперсії показані в рядках 12-15 на рис. 8.6 і 8.7.

Стандартне відхилення варіаційного ряду визначимо як квадратний корінь з дисперсії:

$$\sigma=\sqrt{\sigma^2}=\sqrt{12{,}94}=3{,}6.$$

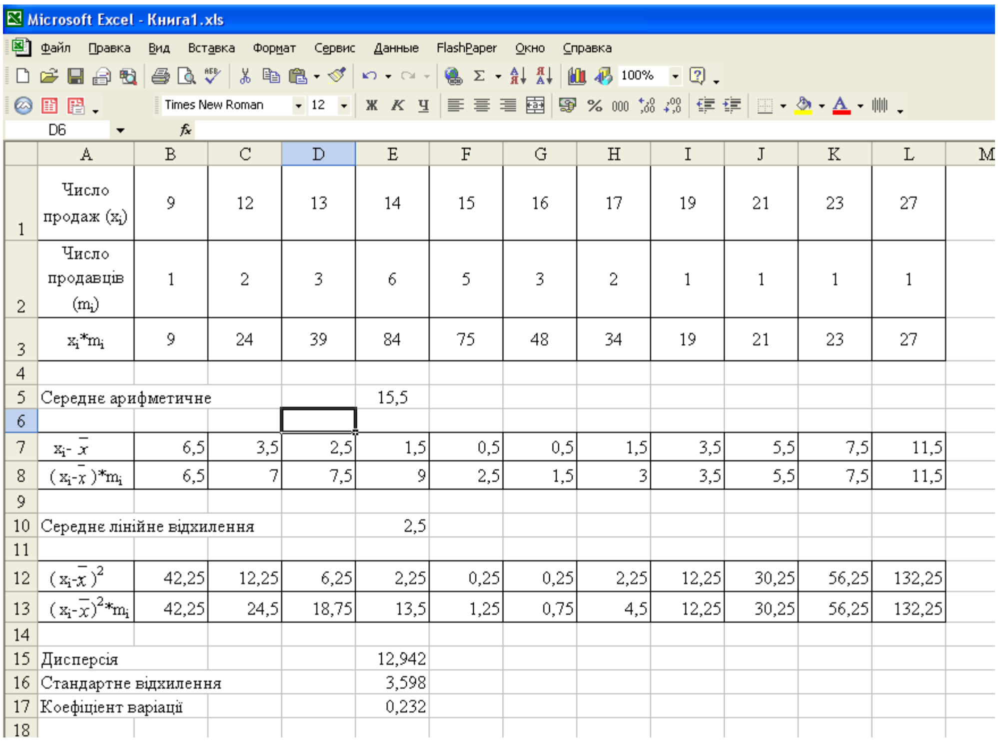
Знайдемо коефіцієнт варіації V як частку стандартного відхилення від середнього значення варіаційного ряду:

$$V=\frac{\sigma}{\bar{x}}=\frac{3{,}6}{15{,}5}=0{,}232.$$



	A	B	C	D	E	F	G	H	I	J
1	Число продаж (x_i)	9	12	13	14	15	16	17	19	21
2	Число продавців (m_i)	1	2	3	6	5	3	2	1	1
3	x_i*m_i	=B1*B2	=C1*C2	=D1*D2	=E1*E2	=F1*F2	=G1*G2	=H1*H2	=I1*I2	=J1*J2
4										
5	Середнє арифметичне				=СУММ(B3:L3)/26					
6										
7	x_i- $\bar{x}$	=ABS(B1-E5)	=ABS(C1-E5)	=ABS(D1-E5)	=ABS(E1-E5)	=ABS(F1-E5)	=ABS(G1-E5)	=ABS(H1-E5)	=ABS(I1-E5)	=ABS(J1-E5)
8	$(x_i-\bar{x})$*m_i	=B7*B2	=C7*C2	=D7*D2	=E7*E2	=F7*F2	=G7*G2	=H7*H2	=I7*I2	=J7*J2
9										
10	Середнє лінійі				=СУММ(B8:L8)/26					
11										
12	$(x_i-\bar{x})^2$	=СТЕПЕНЬ(B7;2)	=СТЕПЕНЬ(C7;2)	=СТЕПЕНЬ(D7;2)	=СТЕПЕНЬ(E7;2)	=СТЕПЕНЬ(F7;2)	=СТЕПЕНЬ(G7;2)	=СТЕПЕНЬ(H7;2)	=СТЕПЕНЬ(I7;2)	=СТЕПЕНЬ(J7;2)
13	$(x_i-\bar{x})^2$*m_i	=B12*B2	=C12*C2	=D12*D2	=E12*E2	=F12*F2	=G12*G2	=H12*H2	=I12*I2	=J12*J2
14										
15	Дисперсія				=СУММ(B13:L13)/					
16	Стандартне ві				=КОРЕНЬ(E15)					
17	Коефіцієнт ва				=E16/E5					
18										

Рис. 8.6 - Використання функцій Microsoft Excel



	A	B	C	D	E	F	G	H	I	J	K	L
1	Число продаж (x_i)	9	12	13	14	15	16	17	19	21	23	27
2	Число продавців (m_i)	1	2	3	6	5	3	2	1	1	1	1
3	x_i*m_i	9	24	39	84	75	48	34	19	21	23	27
4												
5	Середне арифметичне				15,5							
6												
7	$x_i-\bar{x}$	6,5	3,5	2,5	1,5	0,5	0,5	1,5	3,5	5,5	7,5	11,5
8	$(x_i-\bar{x})$*m_i	6,5	7	7,5	9	2,5	1,5	3	3,5	5,5	7,5	11,5
9												
10	Середне лінійне відхилення				2,5							
11												
12	$(x_i-\bar{x})^2$	42,25	12,25	6,25	2,25	0,25	0,25	2,25	12,25	30,25	56,25	132,25
13	$(x_i-\bar{x})^2$*m_i	42,25	24,5	18,75	13,5	1,25	0,75	4,5	12,25	30,25	56,25	132,25
14												
15	Дисперсія				12,942							
16	Стандартне відхилення				3,598							
17	Коефіцієнт варіації				0,232							
18												

Рис. 8.7 - Результати розрахунку числових характеристик варіаційного ряду

Властивості вибіркових числових характеристик

Будь-які значення вибіркових характеристик, обчислені на підставі обмеженого числа елементів вибірки, містять елемент випадковості. Очевидно, що обробка декількох вибірок однакового обсягу дасть ряд різних оцінок відповідної числової характеристики. Отже оцінки числових характеристик є випадковими величинами на відміну від самих числових характеристик, значення яких не випадкові. Необхідно, щоб помилка від заміни дійсного значення числової характеристики його наближеною оцінкою була мінімальною. Такій вимозі задовольняють оцінки числових характеристик, що володіють властивостями спроможності, незміщеності й ефективності.

Оцінка параметра a^* є **спроможною**, якщо при $n\to\infty$ вона збігається за імовірністю до оцінюваного параметра a:

$$\lim_{n\to\infty} a^* = a\,. \tag{8.12}$$

Зокрема, на підставі теореми Чебишева вибіркова середня $\tilde{X}$, визначена за формулами (8.5) і (8.6), є спроможною оцінкою генеральної середньої $\overline{X}$.

Покажемо, що формули (8.9) дають спроможну оцінку генеральної дисперсії. Вибіркова дисперсія визначається як середнє арифметичне квадрата центрованої випадкової величини:

$$\sigma^2=\frac{\sum_{i=1}^{n}(x_i-\tilde{X})^2}{n}=\frac{\sum_{i=1}^{n}x_i^2}{n}+\frac{\sum_{i=1}^{n}\tilde{X}^2}{n}-2\frac{\sum_{i=1}^{n}x_i\tilde{X}}{n}=\alpha_2^*[X]-\left(\tilde{X}\right)^2,$$

де перший доданок – вибірковий другий початковий момент α^*_2; другий доданок – квадрат вибіркової середньої; третій доданок – подвоєний квадрат вибіркової середньої.

У границі дістанемо:

$$\lim_{n\to\infty}\left[\alpha_2^*[X]-\left[\tilde{X}\right]^2\right]=\alpha_2[X]-\left(\overline{X}\right)^2=\sigma_{ген}^2.$$

Таким чином, вибіркова дисперсія, що визначена за формулою (8.9), є спроможною.

Оцінка параметра a^* є **незміщеною** (тобто не містить систематичної помилки), якщо її математичне сподівання дорівнює оцінюваному параметру a:

$$M[a^*]=a. \tag{8.13}$$

Вибіркова середня $\tilde{X}$, визначена за формулами (8.5) і (8.6), є лінійною функцією n незалежних випадкових величин x_i, тому вона сама є випадковою величиною, а отже має свої числові характеристики: математичне сподівання і дисперсію. Покажемо, що математичне сподівання вибіркової середньої не залежить від числа дослідів n і дорівнює генеральній середній:

$$M\left[\tilde{X}\right]=M\left[\frac{\sum x_i}{n}\right]=\frac{1}{n}M\left[\sum x_i\right]=\frac{1}{n}*n*\overline{X}=\overline{X}.$$

Визначимо дисперсію вибіркової середньої:

$$\sigma^2[\tilde{X}]=\sigma^2\left[\frac{\sum x_i}{n}\right]=\left(\frac{1}{n}\right)^2\left(\sum\sigma_i^2\right)=\left(\frac{1}{n}\right)^2 n\sigma^2=\frac{\sigma_{ген}^2}{n}. \tag{8.14}$$

Звідси видно, що чим більше число дослідних даних n, тим менше дисперсія вибіркової середньої, тим вона точніше. З формули (8.14) отримаємо середнє квадратичне відхилення вибіркової середньої:

$$\sigma[\tilde{X}]=\frac{\sigma_{ген}}{\sqrt{n}}. \tag{8.15}$$

Очевидно, що зі збільшенням числа дослідів n $\sigma^2[\tilde{X}]$ прагне до нуля, що свідчить про прагнення $\tilde{X}$ до невипадкової величини $\overline{X}$.

Таким чином, вибіркова середня є незміщеною оцінкою генеральної середньої.

Розглянемо вибіркову дисперсію. Знайдемо її математичне сподівання:

$$M\left[\sigma^2_{выб}\right]=M\left[\alpha^*_2[\,X\,]-\left(\tilde{X}\right)^2\right]=M\left[\frac{\sum\limits_{i=1}^{n}x_i^2}{n}-\left(\frac{\sum\limits_{i=1}^{n}x_i}{n}\right)^2\right]=$$

$$=M\left[\frac{\sum\limits_{i=1}^{n}x_i^2}{n}\right]-M\left[\frac{\sum\limits_{i=1}^{n}x_i^2}{n^2}\right]-2M\left[\frac{\sum\limits_{i<j}x_i\,x_j}{n^2}\right]$$

де перший доданок – математичне сподівання вибіркового другого початкового моменту α^*_2; другий доданок – математичне сподівання вибіркового другого початкового моменту α^*_2, розділеного на n; третій доданок – являє собою вибірковий другий змішаний момент α^*_2, що дорівнює нулю, оскільки значення варіант x_i незалежні, і якщо покласти генеральну середню рівною нулю. Другий і третій доданки отримані при піднесенні у квадрат вибіркової середньої за формулою $(a+b+c)^2 = a^2+b^2+c^2+2ab+2ac+2bc$.

Тоді одержимо, з огляду на те, що генеральна середня дорівнює нулю,

$$M\left[\sigma^2_{выб}\right]=\frac{n}{n}M\left[\frac{\sum\limits_{i=1}^{n}x_i^2}{n}\right]-\frac{1}{n}M\left[\frac{\sum\limits_{i=1}^{n}x_i^2}{n}\right]=\frac{n-1}{n}M\left[\frac{\sum\limits_{i=1}^{n}\left(x_i-\overline{X}\right)^2}{n}\right]=\frac{n-1}{n}\sigma^2_{ген}$$

Таким чином, математичне сподівання вибіркової дисперсії, визначеної за формулою

$$\sigma^2_{выб}=\frac{\sum\left(x_i-\tilde{X}\right)^2}{n} \tag{8.16}$$

не дорівнює генеральній дисперсії, тобто є зміщеною оцінкою:

$$\mathrm{M}[\sigma^2_{выб}]=\frac{n-1}{n}\sigma^2_{ген}.$$

Користуючись цією оцінкою, ми будемо робити систематичну помилку в меншу сторону. Щоб від неї позбутися, необхідно внести виправлення – помножити оцінку дисперсії, отриману за формулою (8.16), на n/(n-1). Незміщена оцінка дисперсії називається також виправленою дисперсією. Її позначають S^2 і знаходять за формулою

$$S^2=\frac{n}{n-1}\sigma^2_{выб}=\frac{\sum\left(x_i-\tilde{X}\right)^2}{n-1} \tag{8.17}$$

Можна показати, що S^2 є незміщеною оцінкою генеральної дисперсії:

$$\mathrm{M}\left[S^2\right]=\frac{n}{n-1}\mathrm{M}\left[\sigma^2_{выб}\right]=\frac{n}{n-1}*\frac{n-1}{n}*\sigma^2_{ген}.$$

Якщо число спостережень n велике, то значення вибіркової дисперсії, обчисленої за формулою (8.16), практично збігається зі значенням, обчисленим за формулою (8.17). При n > 50 вони практично не відрізняються.

Оцінка параметра a^* є **ефективною**, якщо при заданому обсязі вибірки вона має найменшу дисперсію. Ступінь ефективності оцінюють відношенням дисперсій:

$$F = \frac{\sigma_1^2}{\sigma_2^2}, \tag{8.18}$$

тобто якщо F>1, то $\sigma^2{}_2$ більш ефективна, і навпаки.

Якщо на підставі статистичних даних треба визначити оцінки числових характеристик системи двох випадкових величин X і Y, то крім вибіркових середніх і вибіркових дисперсій знаходять оцінку кореляційного моменту за формулою

$$K_{xyвув} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1} \tag{8.19}$$

а оцінку коефіцієнта кореляції - за формулою

$$r_{xyвув} = \frac{K_{xyвув}}{\sigma_{выб_x} \sigma_{выб_y}}. \tag{8.20}$$

Наведені оцінки є спроможними і незміщеними.

Приклад 8.4. Для визначення точності вимірювального приладу було зроблено п'ять незалежних вимірювань, результати яких зведені в таблицю:

Номер вимірювання	1	2	3	4	5
x_i	2781	2836	2807	2763	2858

Визначити незміщену оцінку дисперсії помилок вимірювального приладу, якщо дійсне значення вимірюваної величини:

а) відоме і дорівнює 2800; б) невідоме.

Розв'язання. а) якщо значення вимірюваної величини відоме, то генеральна середня $\bar{X}$ = 2800, незміщена оцінка дисперсії в цьому разі визначиться за формулою (8.16):

$$\sigma^2_{выб} = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{(2781-2800)^2 + (2836-2800)^2 + (2807-2800)^2 + (2763-2800)^2 + (2858-2800)^2}{5} =$$
$$= 1287{,}8$$

б) якщо значення вимірюваної величини невідоме, то слід визначити вибіркову середню, а незміщену оцінку дисперсії обчислити за формулою (8.17)

$$S^2 = \frac{n}{n-1}\sigma^2_{выб} = \frac{\sum (x_i - \tilde{X})^2}{n-1} = \frac{(2781-2809)^2 + (2836-2809)^2 + (2807-2809)^2 + (2763-2809)^2 + (2858-2809)^2}{5-1} =$$
$$= 1508{,}5$$

На рис. 8.8 і 8.9 показаний розрахунок вибіркової дисперсії із застосуванням таблиці Microsoft Excel. У комірці D14 для розв'язання задачі використана функція =ДИСП(B2:F2), що дає незміщену оцінку дисперсії при невідомій генеральній середній. Аргументом функції =ДИСП(B2:F2) є діапазон значень.

	A	B	C	D	E	F	G
1	Номер виміру	1	2	3	4	5	
2	x_i	2781	2836	2807	2763	2858	
3							
4	$(x_i-\bar{x})^2$	361	1296	49	1369	3364	
5							
6	Вибіркова дисперсія			1287,8			
7							
8	Вибіркова середня			2809			
9							
10	$(\tilde{x}_i- \ )^2$	784	729	4	2116	2401	
11							
12	Вибіркова дисперсія			1508,5			
13							
14				1508,5			
15							
16							

Рис. 8.8 - Розрахунок середнього з використанням рядка стану

	A	B	C	D	E	F
1	Номер виміру	1	2	3	4	5
2	x_i	2781	2836	2807	2763	2858
3						
4	$(x_i-\bar{x})^2$	=СТЕПЕНЬ(B2-2800;2)	=СТЕПЕНЬ(C2-2800	=СТЕПЕНЬ(D2-2800	=СТЕПЕНЬ(E2-2800	=СТЕПЕНЬ(F
5						
6	Вибіркова дисперсія			=СУММ(B4:F4)/5		
7						
8	Вибіркова середня			2809		
9						
10	$(x_i-\tilde{x})^2$	=СТЕПЕНЬ(B2-D8;2)	=СТЕПЕНЬ(C2-D	=СТЕПЕНЬ(D2-D	=СТЕПЕНЬ(E2-D	=СТЕПЕНЬ(F
11						
12	Вибіркова дисперсія			=СУММ(B10:F10)/(5		
13						
14				=ДИСП(B2:F2)		

Рис. 8.9 - Розрахунок вибіркової дисперсії із застосуванням таблиці Microsoft Excel

Приклад 8.5. Зроблено вимірювання випадкової величини Y при різних значеннях випадкової величини X. Визначити вибірковий коефіцієнт кореляції цих величин.

x_i	-8	-10	22	2
y_i	-10	-2	4	-1

Розв'язання. Коефіцієнт кореляції визначаємо за формулою

$$r_{xy} = \frac{K_{xy}}{\sigma_x \sigma_y}.$$

Для його обчислення необхідно знайти вибірковий кореляційний момент:

$$K_{xyвув} = \frac{\sum\limits_i (x_i - \overline{X})(y_i - \overline{Y})}{n-1}.$$

Для оцінки середніх значень X і Y знайдемо вибіркові середні:

$$\tilde{X} = \frac{-8+10+22+2}{4} = 6{,}5,$$

$$\tilde{Y} = \frac{-10-2+4-1}{4} = -2{,}25.$$

Визначимо вибіркові дисперсії X і Y:

$$S_x^2 = \frac{(-8-6{,}5)^2 + (10-6{,}5)^2 + (22-6{,}5)^2 + (2-6{,}5)^2}{4-1} = 161;$$

$$S_y^2 = \frac{(-10+2{,}25)^2 + (-2+2{,}25)^2 + (4+2{,}25)^2 + (-1+2{,}25)^2}{4-1} = 33{,}6.$$

Визначимо вибіркові середні квадратичні відхилення:

$$\sigma_{выб_x} = \sqrt{S_x^2} = \sqrt{161} = 12{,}7 \qquad \sigma_{выб_y} = \sqrt{S_y^2} = \sqrt{33{,}6} = 5{,}8.$$

Тепер можна розрахувати кореляційний момент

$$K_{xyвув} = \frac{(-8-6{,}5)(-10+2{,}25) + (10-6{,}5)(-2+2{,}25) + (22-6{,}5)(4+2{,}25) + (2-6{,}5)(-1+2{,}25)}{4-1} = 68{,}2$$

і коефіцієнт кореляції

$$r_{xy} = \frac{K_{xyвув}}{\sigma_{выб_x}\sigma_{выб_y}} = \frac{68{,}2}{12{,}7*5{,}8} = 0{,}926.$$

Довірчий інтервал і довірча імовірність

Розглянуті оцінки параметрів є точковими, тому що характеризуються одним числом. Якщо точкова оцінка параметра визначена на підставі вибірки малого обсягу, вона може значно відрізнятися від оцінюваного параметра. Для визначення помилки від заміни генерального параметра його оцінкою використовують поняття довірчого інтервалу і довірчої імовірності.

Нехай для деякого параметра розподілу, наприклад, математичного сподівання m_x отримано спроможну і незміщену оцінку a*. Треба знати, особливо при малому числі спостережень, до яких помилок може призвести заміна пара метра m_x його точковою оцінкою, і з яким ступенем впевненості можна чекати, що ці помилки не вийдуть за певні межі. Візьмемо досить велику імовірність β (0,95; 0,99) таку, що подію $A=\{|a^*-m_x|<l\}$, яка характеризується цією імовірністю, можна вважати практично достовірною, і знайдемо таке значення l, для якого справедлива рівність

$$P(A) = P\{(a-l)< m_x <(a+l)\} = \beta. \quad (8.21)$$

Рівність (8.21) означає, що діапазон можливої помилки при заміні m_x на a^* з імовірністю β дорівнюватиме $\pm$ ***l***. Тут a^* і β відомі, ***l*** підлягає визначенню. З імовірністю β невідоме значення параметра m_x буде перебувати в інтервалі $L=[a^*-l,\ a^*+l]$. Більші за абсолютним значенням помилки, ніж l, будуть зустрічатися з імовірністю

$$\alpha = 1 - \beta.$$

Межі інтервалу називаються довірчими межами:

$$a_1 = a^*-l; \qquad a_2 = a^*+l.$$

Зазначимо, що величина m_x невипадкова, а випадковим є положення відрізка L=2*l на осі абсцис, обумовлене випадковою величиною a^*. Таким чином, довірчу імовірність β можна трактувати як імовірність того, що випадковий інтервал L накриє точку m_x (рис. 8.10).

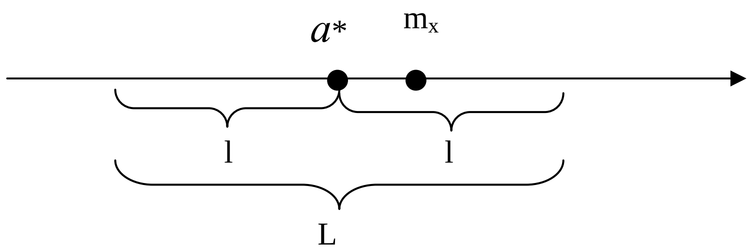


Рис. 8.10 - Довірчий інтервал

Таким чином, значення β і L характеризують ступінь впевненості й величину похибки при встановленні дійсного значення шуканого параметра m_x.
Нехай для параметра a є незміщена оцінка a^*. Якби нам був відомий закон розподілу величини a^*, то задача знаходження довірчого інтервалу була б проста: досить знайти таке значення l, для якого

$$P\{|a^*-a| < l\} = \beta.$$

Утруднення полягає в тому, що закон розподілу a^* залежить від закону розподілу досліджуваної ознаки X і, отже від його невідомих параметрів (зокрема і від самого параметра a). Щоб обійти це утруднення, застосовують наступний грубо наближений прийом: заміняють у виразі для l невідомі параметри їхніми точковими значеннями. При 20-30 дослідах цей прийом звичайно дає задовільні за точністю результати.

Нехай зроблено n незалежних дослідів, з яких знайдені значення вибіркової середньої $\tilde{X}$ і вибіркової дисперсії S^2. Потрібно визначити довірчий інтервал для вибіркової середньої $\tilde{X}$.

Оскільки вибіркова середня визначається за формулою

$$\tilde{X} = \frac{\sum_{i=1}^{n} x_i}{n},$$

її можна розглядати як функцію n незалежних випадкових величин x_i, що в загальному випадку можуть бути підпорядковані будь-яким законам розподілу. Тоді відповідно до центральної граничної теореми закон розподілу вибіркової середньої $\tilde{X}$ буде близький до нормального з параметрами

$$M\left[\tilde{X}\right] = \bar{X};$$

$$\sigma^2\left[\tilde{X}\right] = \sigma^2\left[\frac{\sum_{i=1}^{n} x_i}{n}\right] = \frac{1}{n^2}\sum \sigma^2[x_i] = \frac{n\sigma^2_{ген}}{n^2} = \frac{\sigma^2_{ген}}{n};$$

$$\sigma\left[\tilde{X}\right] = \frac{\sigma_{ген}}{\sqrt{n}}.$$

Припустимо, що генеральна дисперсія відома, тоді скориставшись інтегралом імовірностей

$$\Phi(x) = \frac{1}{\sqrt{\pi}}\int_0^x e^{-\frac{t^2}{2}}dt,$$

і врахувавши, що помилка симетрична відносно a^*, а β задана, дістанемо:

$$P\{(a^* - l) < \tilde{X} < (a^* + l)\} = \beta = 2\Phi\left(\frac{l}{\sigma[\tilde{X}]}\right).$$

Користуючись таблицею значень інтеграла імовірностей, можемо знайти аргумент

$$\frac{l}{\sigma[\tilde{X}]} = \Phi^{-1}(\beta), \text{ звідси } \quad l = \sigma\left[\tilde{X}\right]\Phi^{-1}(\beta).$$

Довірчий інтервал L = 2*l.

Приклад 8.6. Знайдемо довірчий інтервал для вибіркової середньої досліджуваної ознаки X. Нехай зроблено n незалежних дослідів і визначені спроможні й незміщені оцінки параметрів цієї ознаки $\tilde{X}$ і $\sigma^2_{виб}$.

Розв'язання. Нехай $\tilde{X}$ = 10, $\sigma^2_{виб}$ = 4, n = 40. Задамося значенням довірчої імовірності β = 0,95. Тоді можна записати

$$P\{(\tilde{X} - l) < \bar{X} < (\tilde{X} + l)\} = 0{,}95.$$

Скористаємося тим, що випадкова величина $\tilde{X}$ є функцією n незалежних випадкових величин x_i. Тоді відповідно до центральної граничної теореми щільність розподілу випадкової величини $\tilde{X}$ практично буде підпорядковуватися нормальному законові розподілу з параметрами

$$M[\tilde{X}] = \bar{X} = 10, \; D[\tilde{X}] = \frac{\sigma^2_{ген}}{n} = 4/40 = 0{,}1.$$

Для нормального закону розподілу імовірність влучення випадкової величини на інтервал значень можна виразити за допомогою інтеграла імовірностей

$$P\{(m_x^* - l) \le m_x \le (m_x^* + l)\} = \left[\Phi\left(\frac{m_x^* + l - m_x}{\sqrt{\ }\sigma_x} \right) - \Phi\left(\frac{m_x^* - l - m_x}{\sqrt{\ }\sigma_x} \right) \right] =$$

$$= \left[\Phi\left(\frac{l}{\sqrt{\ }\sigma_x} \right) - \Phi\left(\frac{-l}{\sqrt{\ }\sigma_x} \right) \right] = 2\Phi\left(\frac{l}{\sqrt{\ }\sigma_x} \right) = \beta .$$

Підставимо значення:

$$\Phi\left(\frac{l}{\sqrt{0,32}} \right) = 0,475; \qquad \frac{l}{0,453} = 1,4,$$

звідки l = 1,4 * 0,453 = 0,634.

Таким чином, з імовірністю 0,95 інтервал (9,366; 10,634) накриє генеральну середню досліджуваної ознаки X.

Приклад 8.7. Результати вимірювання значень випадкових величин X і Y наведені в таблиці:

№ досліду	x_i	y_i
1	4	0,041
2	8	0,05
3	10	0,081
4	14	0,104
5	16	0,12
6	20	0,139
7	19	0,154
8	23	0,18
9	26	0,208
10	30	0,241
11	31	0,25
12	36	0,269
13	37	0,301

Визначити точкові й інтервальні оцінки числових характеристик системи випадкових величин X і Y, а також імовірність того, що вибіркова середня випадкової величини X відрізняється від її генеральної середньої не більше ніж на 1.

Розв'язання. Проміжні розрахунки будемо зводити в таблицю. Спочатку визначимо точкові значення вибіркових середніх:

$$\tilde{Y} = \frac{\sum y_i}{n} = \frac{2,138}{13} = 0,164;$$

$$\tilde{X} = \frac{\sum x_i}{n} = \frac{274}{13} = 21,1.$$

№ досліду	x_i	y_i	$\overset{o}{x}_i$	$\overset{o}{x}{}^2$	$\overset{o}{y}_i$	$\overset{o}{y}{}^2$	$\overset{o}{x}_i\,\overset{o}{y}_i$
1	4	0,041	-17,1	292,41	-0,123	0,0151	2,103
2	8	0,05	-13,1	171,61	-0,114	0,0130	1,493
3	10	0,081	-11,1	123,21	-0,083	0,0069	0,921
4	14	0,104	-7,1	50,41	-0,06	0,0036	0,426
5	16	0,12	-5,1	26,01	-0,044	0,0019	0,224
6	20	0,139	-1,1	1,21	-0,025	0,0006	0,028
7	19	0,154	-2,1	4,41	-0,01	0,0001	0,021
8	23	0,18	1,9	3,61	0,016	0,0003	0,030
9	26	0,208	4,9	24,01	0,044	0,0019	0,216
10	30	0,241	8,9	79,21	0,077	0,0059	0,685
11	31	0,25	9,9	98,01	0,086	0,0074	0,851
12	36	0,269	14,9	222,01	0,105	0,0110	1,565
13	37	0,301	15,9	252,81	0,137	0,0188	2,178
Сума	$\Sigma x_i=274$	$\Sigma y_i=2{,}138$		$\sum \overset{o}{x}{}_i^2$		$\sum \overset{o}{y}{}_i^2$	$\sum \overset{o}{x}_i\,\overset{o}{y}_i$

Знайдемо точкові значення незміщених вибіркових дисперсій:

$$S_y^2=\frac{\sum\limits_1^{13}(y_i-\tilde{Y})^2}{13-1}=\frac{\sum\limits_1^{13}\overset{o}{y}{}^2}{12}=\frac{0{,}0866}{12}=0{,}0072;$$

$$S_x^2=\frac{\sum\limits_1^{13}(x_i-\tilde{X})^2}{13-1}=\frac{\sum\limits_1^{13}\overset{o}{x}{}^2}{12}=\frac{1348{,}93}{12}=112{,}41.$$

Визначимо вибіркові середні квадратичні відхилення:

$$\sigma_{\text{выб}_y}=\sqrt{0{,}0072}=0{,}085;\quad \sigma_{\text{выб}_x}=\sqrt{112{,}41}=10{,}6.$$

Знайдемо вибірковий кореляційний момент за формулою (8.19)

$$K_{xy\text{выб}}=\frac{\sum\limits_1^{13}\overset{o}{x}\,\overset{o}{y}}{13-1}=\frac{10{,}742}{12}=0{,}895$$

і коефіцієнт кореляції за формулою (8.20)

$$r_{xy\text{выб}}=\frac{K_{xy\text{вув}}}{\sigma_{\text{выб}_y}*\sigma_{\text{выб}_x}}=\frac{0{,}895}{0{,}085*10{,}6}=0{,}994.$$

Для визначення імовірності того, що помилка від заміни генеральної середньої випадкової величини X її вибірковою оцінкою не перевершить 1, побудуємо довірчий інтервал з межами 21,1±1 і знайдемо імовірність того, що цей інтервал накриє генеральну середню випадкової величини X (довірчу імовірність β).

Скористаємося тим, що величина $\tilde{X}$ являє собою суму n=13 незалежних однаково розподілених випадкових величин x_i і, відповідно до центральної граничної теореми, при досить великому n її закон розподілу близький до нормального (а в нашім випадку ми провадили вимірювання, що завжди дає помилку, розподілену за нормальним законом). Будемо виходити з того, що величина $\tilde{X}$ розподілена за нормальним законом з характеристиками:

$$M[\tilde{X}] = \overline{X} = 21{,}1;$$

$$\sigma[\tilde{X}] = \sqrt{\frac{\sigma^2_{ген}}{n}} = \frac{10{,}6}{\sqrt{13}} = 2{,}94.$$

Виразимо шукану імовірність β за допомогою функції Лапласа:

$$P\left\{\left|\tilde{X} - \overline{X}\right| < l\right\} = \beta = 2\Phi\left(\frac{l}{\sigma[\tilde{X}]}\right),$$

звідки дістанемо

$$2\Phi\left(\frac{1}{2{,}94}\right) = \beta;\ \beta = 2\Phi(0{,}34) = 0{,}266.$$

Отримане значення імовірності дуже мале, тобто подія, яка полягає в тому, що помилка від заміни генеральної середньої X її вибірковою оцінкою не перевершить 1, практично неможлива. Задамося довірчою імовірністю β =0,95 і визначимо межі відповідного їй довірчого інтервалу. Для цього запишемо вираження імовірності β

$$2\Phi\left(\frac{l}{2{,}94}\right) = \beta = 0{,}95;\ \Phi\left(\frac{l}{2{,}94}\right) = 0{,}95/2 = 0{,}475,$$

звідки, користуючись таблицею значень інтеграла імовірностей, одержимо

$$\frac{l}{2{,}94} = 1{,}96.$$

Таким чином, ***l***=1,96*2,94 = 5,76. З імовірністю 0,95 інтервал (15,34; 26,86) накриє генеральну середню випадкової величини X. Помилка становить 27,3%, тобто занадто велика. Для зменшення помилки треба збільшити число вимірювань.

Побудуємо довірчий інтервал для вибіркової дисперсії випадкової величини X. Закон розподілу вибіркової дисперсії також наближається до нормального. Один з параметрів розподілу - математичне сподівання вибіркової дисперсії

$$M[S^2_x] = \sigma^2_{ген.}.$$

Дисперсію вибіркової дисперсії S^2_x обчислимо за формулою

$$D[S^2_x] = \frac{1}{n}\mu_4 - \frac{n-3}{n(n-1)}\sigma^2_{ген},$$

де μ_4 – центральний момент випадкової величини X четвертого порядку. Його оцінку можна обчислити за формулою

$$\mu_4^* = \frac{1}{n}\sum_i (x_i - \tilde{X})^4 .$$

Але справа в тому, що при обмеженому числі дослідів моменти високого порядку визначаються з великими помилками, тому якщо величина X розподілена за нормальним законом, то μ_4 можна обчислити через дисперсію:

$$\mu_4 = 3 * \sigma^2_{ген} .$$

Отримаємо:

$$D[S_x^2] = \frac{3}{n}\sigma^2_{ген} - \frac{n-3}{n(n-1)}\sigma^2_{ген} = \frac{3}{13}(114{,}041)^2 - \frac{13-3}{13(13-1)}(114{,}041)^2 = 2106 ,$$

звідси

$$\sigma[S_x^2] = \sqrt{2106} = 45{,}9 .$$

Визначимо довірчий інтервал для вибіркової дисперсії, задавшись імовірністю $\beta = 0{,}9$.

Користуючись тим, що S^2_x розподілена нормально, виразимо імовірність β за допомогою функції Лапласа:

$$P\left\{\left|S_x^2 - D_x\right| < l\right\} = \beta = 2\Phi\left(\frac{l}{\sigma[S_x^2]}\right), \text{ звідки дістанемо } \Phi\left(\frac{l}{45{,}9}\right) = 0{,}45$$

$$\frac{l}{45{,}9} = 1{,}64 , \qquad l = 1{,}64 * 45{,}9 = 75{,}26.$$

Таким чином, з імовірністю 0,9 дисперсія X лежить в інтервалі 112,41±75,26. Помилка становить 67% - дуже велика.

Приклад 8.8. У попередньому прикладі результати обчислення показали, що для зменшення довірчого інтервалу (15,34; 26,86), який з довірчою імовірністю 0,95 накриє генеральну середню випадкової величини X, необхідно збільшити число вимірювань. Визначимо, яким повинне бути число вимірювань n, щоб помилка від заміни генеральної середньої випадкової величини X її вибірковою оцінкою $\tilde{X} = 21{,}1$ не перевершувала 1 з імовірністю β=0,95.

Розв'язання: Виразимо довірчу імовірність β за допомогою функції Лапласа:

$$P\left\{\left|\tilde{X} - \overline{X}\right| < l\right\} = \beta = 2\Phi\left(\frac{l}{\sigma[\tilde{X}]}\right),$$

звідки дістанемо, підставивши l=1 і β =0,95,

$$\Phi\left(\frac{1}{\sigma[\tilde{X}]}\right) = 0{,}475 ,$$

тоді $\frac{1}{\sigma[\tilde{X}]} = 1{,}96$, звідки $\sigma[\tilde{X}] = \frac{1}{1{,}96} = 0{,}510$. Відомо, що середнє квадратичне відхилення вибіркової середньої $\sigma[\tilde{X}] = \sqrt{\frac{\sigma^2_{ген}}{n}}$. Знайдемо необхідне число вимірів n у такий спосіб:

$$n = \left(\frac{\sigma^2_{ген}}{\sigma[\,\tilde{X}\,]} \right)^2 = \left(\frac{10{,}6^2}{0{,}510} \right)^2 = 432\,.$$

Таким чином, для досягнення необхідної точності вибіркової середньої замість 13 дослідів треба зробити 432 вимірювання випадкової величини X.

Приклад 8.9. Результати вимірювання значень випадкових величин U і I наведені в таблиці:

Номер досліду	Ii	Ui	Номер досліду	Ii	Ui
1	-18	-30	11	4	-8
2	-16	-32	12	6	22
3	-14	-10	13	8	-2
4	-12	-8	14	10	10
5	-10	-15	15	12	3
6	-8	10	16	14	18
7	-6	5	17	16	55
8	-4	2	18	18	28
9	-2	3	19	20	22
10	2	8	20	22	62

Знайти оцінки числових характеристик системи випадкових величин U і I і визначити 80-відсотковий довірчий інтервал для вибіркових середніх.

Розв'язання. Для розв'язання задачі скористаємося статистичними функціями Microsoft Excel. Незміщені спроможні вибіркові середні визначимо за допомогою функції СРЗНАЧ. Функція СРЗНАЧ повертає середнє арифметичне своїх аргументів. Синтаксис функції =СРЗНАЧ(число1, число2,...), де число1, число2, - це від 1 до 30 аргументів, для яких обчислюється середнє. Аргументи повинні бути або числами, або іменами, масивами або посиланнями, що містять числа. Якщо аргумент, який є масивом або посиланням, містить тексти, логічні значення або порожні комірки, то такі значення ігноруються; однак комірки, які містять нульові значення, враховуються. В якості аргументу функції вкажемо для випадкової величини I масив комірок B2:B21, для випадкової величини U - масив комірок C2:C21 (рис. 8.12) і отримаємо

$$\tilde{I} = \frac{\sum i_i}{n} = 2{,}1; \qquad \tilde{U} = \frac{\sum u_i}{n} = 7{,}15\,.$$

Для обчислення оцінок вибіркових дисперсій скористаємося функцією ДИСП. Функція ДИСП оцінює дисперсію за вибіркою. Синтаксис функції =ДИСП(число1;число2; ...), де число1, число2, ... - це від 1 до 30 числових аргументів, що відповідають вибірці з генеральної сукупності. Припускаємо, що аргументи є тільки вибіркою з генеральної сукупності. Якщо дані становлять всю генеральну сукупність, для обчислення дисперсії слід використовувати іншу функцію. Логічні значення, такі як ИСТИНА або ЛОЖЬ, а також текст ігноруються. Функція ДИСП використовує наступну формулу:

$$S^2=\frac{n\sum_1^n x_i^{\,2}-\left(\sum_1^n x_i\right)^2}{n(n-1)}.$$

Як аргумент функції вкажемо для випадкової величини I масив комірок B2:B21, для випадкової величини U - масив комірок C2:C21 (рис. 8.12) і отримаємо

$$S_I^2=\frac{n\sum_1^{20} i_i^2-(\sum_1^{20} i_i)^2}{n(n-1)}=161{,}88; \qquad S_u^2=\frac{n\sum_1^{20} u_i^2-(\sum_1^{20} u_i)^2}{n(n-1)}=558{,}66.$$

Для обчислення середніх квадратичних відхилень є кілька статистичних функцій. Скористаємося функцією СТАНДОТКЛОН. Функція СТАНДОТКЛОН оцінює стандартне відхилення за вибіркою, тобто передбачає, що аргументи є тільки вибіркою з генеральної сукупності. Якщо дані являють всю генеральну сукупність, то стандартне відхилення необхідно обчислювати за допомогою іншої функції. Стандартне відхилення обчислюється з використанням «незміщеного» методу, функція СТАНДОТКЛОН використовує наступну формулу:

$$\sigma=\sqrt{\frac{n\sum_1^{20} x_i-(\sum x_i)^2}{n(n-1)}}.$$

Логічні значення, такі як ИСТИНА або ЛОЖЬ, а також текст ігноруються. Дамо посилання на масиви B2:B21 і C2:C21 і отримаємо

$$\sigma_{выб_i}=12{,}72, \qquad \sigma_{выб_U}=23{,}64.$$

Кореляційний момент (коваріацію) розрахуємо з використанням функції КОВАР. Функція КОВАР повертає коваріацію, тобто середнє добутків відхилень для кожної пари точок даних. Коваріація використовується для визначення зв'язку між двома множинами даних. Синтаксис функції =КОВАР(масив1; масив2), де масив1 - це перший масив або інтервал даних, масив2 - це другий масив або інтервал даних. Аргументи повинні бути числами або іменами, масивами або посиланнями, що містять числа. Якщо аргумент, який є масивом або посиланням, містить тексти, логічні значення або порожні комірки, то такі значення ігноруються, однак комірки, які містять нульові значення, враховуються. Якщо масив1 і масив2 мають різне число даних, то КОВАР повертає значення помилки #Н/Д. Якщо масив1, або масив2 порожні, то КОВАР повертає значення помилки #ДЕЛ/0!.

Коваріація визначається в такий спосіб:

$$Cov(X,Y)=\frac{\sum_1^{20}(x_i-\tilde{x})(y_i-\tilde{y})}{n},$$

обчислимо кореляційний момент (рис 8.11):

$$K_{iu} = Cov(I,U) = 234{,}49$$

Для обчислення коефіцієнта кореляції скористаємось функцією КОРРЕЛ. Функція КОРРЕЛ повертає коефіцієнт кореляції між інтервалами комірок масив1 і масив2. Синтаксис: =КОРРЕЛ(масив1;масив2), де масив1 і масив2- це комірки інтервалів значень. Аргументи повинні бути числами або іменами, масивами або посиланнями, що містять числа. Якщо аргумент, що є масивом або посиланням, містить текст, логічні значення або порожні комірки, то такі значення ігноруються; однак комірки, які містять нульові значення, враховуються. Якщо масив1 і масив2 мають різну кількість точок даних, то функція КОРРЕЛ повертає значення помилки #Н/Д. Якщо масив1 або масив2 порожній, або якщо σ (стандартне відхилення) їх значень дорівнює нулю, то функція КОРРЕЛ повертає значення помилки #ДЕЛ/0!. Рівняння для коефіцієнта кореляції має такий вигляд:

$$r_{iu} = \frac{Cov(I,U)}{\sigma_i \sigma_u} = 0{,}82 .$$

	A	B	C	D	E
1	Номер досліду	Ii	Ui		
2	1	-18	-30	Вибіркова середня I =	=СРЗНАЧ(B2:B21)
3	2	-16	-32	Вибіркова середня U =	=СРЗНАЧ(C2:C21)
4	3	-14	-10	Вибіркова дисперсія I =	=ДИСП(B2:B21)
5	4	-12	-8	Вибіркова дисперсія U =	=ДИСП(C2:C21)
6	5	-10	-15	Середнє квадратичне відхилення I =	=СТАНДОТКЛОН(B2:B21)
7	6	-8	10	Середнє квадратичне відхилення U =	=СТАНДОТКЛОН(C2:C21)
8	7	-6	5	Кореляційний момент Kiu =	=КОВАР(B2:B21;C2:C21)
9	8	-4	2	Коефіцієнт кореляції r_{iu} =	=КОРРЕЛ(B2:B21;C2:C21)
10	9	-2	3		
11	10	2	8		
12	11	4	-8	довірчий інтервал для вибіркової середньої I	=ДОВЕРИТ(0,2;E6;20)
13	12	6	22		
14	13	8	-2	довірчий інтервал для вибіркової середньої U	=ДОВЕРИТ(0,2;E7;20)
15	14	10	10		=E2-E12
16	15	12	3	межі довірчого інтервалу для вибіркової середньої I	=E2+E12
17	16	14	18		
18	17	16	55		=E3-E14
19	18	18	28	межі довірчого інтервалу для вибіркової середньої U	=E3+E14
20	19	20	22		
21	20	22	62		

Рис. 8.11 - Обчислення кореляційного моменту

Тепер визначимо довірчий інтервал для $\tilde{I}$ і $\tilde{U}$ при β = 0,8. Скористаємось функцією ДОВЕРИТ. Функція ДОВЕРИТ повертає довірчий інтервал для середньої генеральної сукупності. Синтаксис функції = ДОВЕРИТ (альфа;станд_откл;размер), де альфа - рівень значущості, використовуваний для обчислення рівня надійності; станд_откл - стандартне відхилення генеральної сукупності для інтервалу даних, передбачається відомим; размер - це розмір вибірки. Якщо який-небудь з аргументів не є числом, то функція ДОВЕРИТ повертає значення помилки #ЗНАЧ!. Якщо альфа ≤ 0 або альфа ≥ 1, то функція ДО-

ВЕРИТ повертає значення помилки #ЧИСЛО!. Якщо станд_откл ≤ 0, то функція ДОВЕРИТ повертає значення помилки #ЧИСЛО!. Якщо розмір не ціле число, то воно усікається до цілого. Якщо розмір < 1, то функція ДОВЕРИТ повертає значення помилки #ЧИСЛО!

Дістанемо значення 80-процентних довірчих інтервалів:

для I $l_i = 3{,}65;$

для U $l_u = 6{,}77.$

Тоді межі довірчих інтервалів

для I

$$\tilde{I}_1 = 2{,}1-3{,}65= -1{,}55 \quad \tilde{I}_2 = 2{,}1+3{,}65= 5{,}75.$$

для U

$$\tilde{U}_1 = 7{,}15-6{,}77= 0{,}38 \quad \tilde{U}_2 = 7{,}15+6{,}77= 13{,}92.$$

Результати розрахунків показані на рис. 8.12.

	A	B	C	D	E	F
1	Номер досліду	Ii	Ui			
2	1	-18	-30	Вибіркова середня I =	2,1	
3	2	-16	-32	Вибіркова середня U =	7,15	
4	3	-14	-10	Вибіркова дисперсія I =	161,88	
5	4	-12	-8	Вибіркова дисперсія U =	558,66	
6	5	-10	-15	Середнє квадратичне відхі	12,72	
7	6	-8	10	Середнє квадратичне відхі	23,64	
8	7	-6	5	Кореляційний момент Kiu	234,49	
9	8	-4	2	Коефіцієнт кореляції r_{iu} =	0,82	
10	9	-2	3			
11	10	2	8	довірчий інтервал для		
12	11	4	-8	вибіркової середньої I	3,65	
13	12	6	22	довірчий інтервал для		
14	13	8	-2	вибіркової середньої U	6,77	
15	14	10	10	межі довірчого інтервалу	-1,55	
16	15	12	3	для вибіркової середньої I	5,75	
17	16	14	18			
18	17	16	55	межі довірчого інтервалу	0,38	
19	18	18	28	для вибіркової середньої	13,92	
20	19	20	22			
21	20	22	62			

Рис. 8.12 - Визначення меж довірчого інтервалу

Приклад 8.10. Зроблено 20 дослідів над досліджуваною ознакою X, результати яких наведені в таблиці

i	x_i	i	x_i
1	10,5	11	10,6
2	10,8	12	11,3
3	11,2	13	10,5
4	10,9	14	10,7
5	10,4	15	10,8
6	10,6	16	10,9
7	10,9	17	10,8
8	11,0	18	10,7
9	10,3	19	10,9
10	10,8	20	11

Знайти точкову й інтервальну оцінки вибіркової середньої.

Розв'язання. Для визначення точкової і інтервальної оцінок вибіркової середньої досліджуваної ознаки X скористаємося функціями Microsoft Excel СРЗНАЧ, СТАНДОТКЛОН і ДОВЕРИТ (рис. 8.13). Попередньо задамося довірчою імовірністю β=0,95.

	A	B	C	D	E
1	i	x	$\overset{\circ}{x}$		
2	1	10,5	-0,28		
3	2	10,8	0,02	Вибіркова середня	10,78
4	3	11,2	0,42	Стандартне відхилення	0,25
5	4	10,9	0,12		
6	5	10,4	-0,38	Довірчий інтервал	0,111
7	6	10,6	-0,18		
8	7	10,9	0,12		
9	8	11	0,22		
10	9	10,3	-0,48		
11	10	10,8	0,02		
12	11	10,6	-0,18		
13	12	11,3	0,52		
14	13	10,5	-0,28		
15	14	10,7	-0,08		
16	15	10,8	0,02		
17	16	10,9	0,12		
18	17	10,8	0,02		
19	18	10,7	-0,08		
20	19	10,9	0,12		
21	20	11	0,22		

Рис. 8.13 - Визначення точкової і інтервальної оцінок вибіркової середньої

Таким чином, вибіркова середня дорівнює 10,78 і з імовірністю 0,95 інтервал (10,67; 10,89) містить генеральну середню. Або, інакше кажучи, користуючись вибірковою середньою 10,78 замість генеральної середньої, ми з 95-відсотковою впевненістю робимо помилку, що не перевершує ±0,111 (або 1%).

Запитання для самоперевірки:

1. Які завдання вирішує математична статистика? Назвіть основні з них.
2. Поясніть зміст вибіркового методу.
3. У чому полягає різниця між генеральною сукупністю і вибіркою?
4. Яку інформацію про досліджувану ознаку дістають з варіаційного ряду?
5. Що таке оцінка параметра розподілу?
6. Якими властивостями повинні володіти вибіркові числові характеристики варіаційного ряду?
7. Поясніть властивості спроможності й незмещеності оцінок.
8. Чим відрізняються точкова і інтервальна оцінки параметрів розподілу?
9. Поясніть поняття «довірчий інтервал» і «довірча імовірність».

Задачі для самостійного розв'язання

8.1. На біржі протягом деякого часу проводяться статистичні дослідження коливань ціни на партії товарів. Результати спостережень: 3100, 4000, 3800, 4100, 3400, 4200, 3700, 3900, 3200, 4100, 3800, 4200, 3500, 4000, 3900. Побудувати варіаційний ряд і гістограму випадкової величини X - ціна на партію товарів.

8.2. Для умов попередньої задачі обчислити числові характеристики варіаційного ряду.

8.3. Для умов задач 8.1 і 8.2 визначити довірчі інтервали для вибіркової середньої з надійністю 0,95 і вибіркової дисперсії з надійністю 0,99. Прийняти, що досліджувана ознака X розподілена нормально.

Тема 9. ЕЛЕМЕНТИ ТЕОРІЇ КОРЕЛЯЦІЇ

Завданням кореляційного аналізу є визначення форми залежності між незалежною і залежною змінними X і Y і оцінка тісноти зв'язку між ними.

Залежність виду

$$y = f(x), \tag{9.1}$$

в якій кожному значенню X відповідає одне певне значення Y, називається **функціональною**.

Одному значенню змінної X x_i може відповідати ряд значень змінної Y: $y_{i1}, y_{i2}, \dots, y_{ik}$, що може бути викликане впливом різних факторів на змінну Y, а не тільки фактора X, або помилками вимірювання. У цьому випадку залежність називається **статистичною.** Для кожного значення x_i можна визначити умовне середнє $\bar{y}_i$ (рис. 9.1).

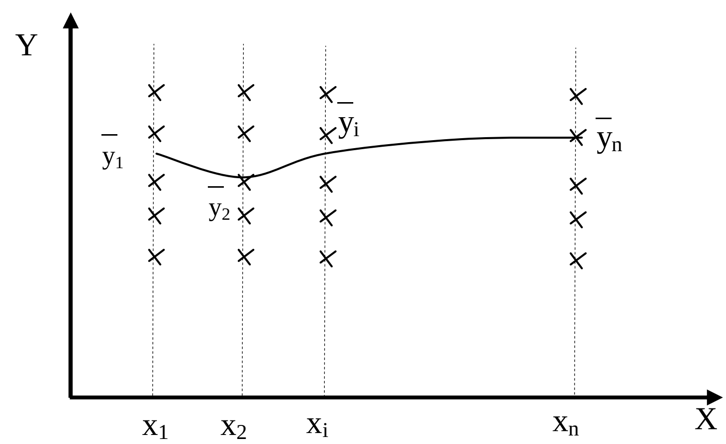


Рис. 9.1 - Статистична залежність y = f(x)

Статистична залежність між змінними Y і X передбачає, що при зміні незалежної змінної X змінюється розподіл залежної змінної Y. Якщо при зміні змінної X закон розподілу Y не змінюється, а змінюється її середнє значення, то така статистична залежність називається **кореляційною**:

$$\bar{y}_x = \varphi(x). \tag{9.2}$$

Кореляційний аналіз заснований на використанні рівняння регресії.

Регресією Y на X називається умовне математичне сподівання випадкової величини Y за умови, що X прийняла значення x_i. Лінія, що з'єднує точки $\bar{y}_i$, називається **лінією регресії** (рис. 9.1).

Для апроксимації лінії регресії аналітичним виразом використовують **рівняння регресії** (9.2). Незалежна змінна X називається пояснюючою змінною або ознакою-фактором. Залежна змінна Y називається пояснювальною ознакою або результативною ознакою. Розрізняють парну (або просту) регресію, якщо досліджують вплив на результативну ознаку Y одного фактора X, і множинну регресію, якщо досліджують вплив на результативну ознаку Y множини факторів X_i.

Вибір вигляду залежності $\overline{y}_x = \varphi(x)$ може бути здійснений з теоретичних міркувань або графічно, для чого залежність зображують точками на координатній площині. Таке зображення статистичної залежності називається **полем кореляції.** Наприклад, розташування отриманих точок на рис. 9.2. нагадує параболу, тоді для згладжування експериментальної залежності $\overline{y}_x = \varphi(x)$ можна скористатися поліномом другого порядку:

$$y = a_0 + a_1x + a_2x^2.$$

На практиці найчастіше використовують лінійне рівняння регресії:

$$Y = \rho_{yx} X + b \qquad (9.3)$$

Коефіцієнт при змінній X ρ_{yx} називається коефіцієнтом регресії.

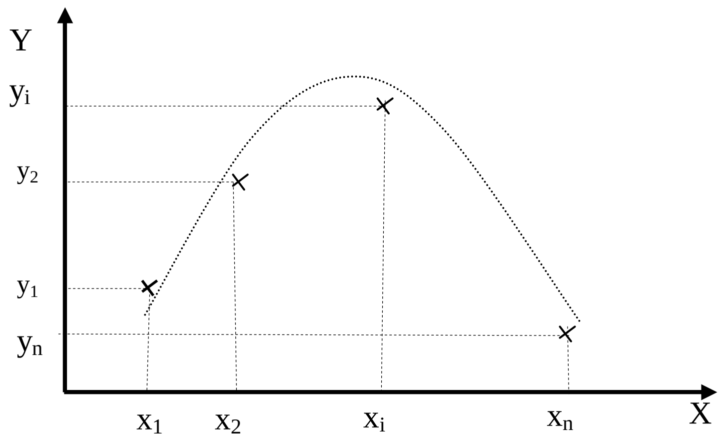


Рис. 9.2 - Поле кореляції

Метод найменших квадратів

Для визначення параметрів залежності, що згладжує $\overline{y}_x = \varphi(x)$, зокрема значень параметрів ρ_{yx} і b рівняння регресії (9.3), застосовують ***метод найменших квадратів*** (МНК). Метод найменших квадратів дозволяє при відомому класі апроксимуючої залежності $\overline{y}_x = \varphi(x)$ так вибрати значення її параметрів, щоб ця залежність щонайкраще відображала дані спостережень.

Нехай в результаті n дослідів для кожного значення фактора $X=(x_1, x_2, \ldots, x_n)$ отримані значення результативної ознаки $Y=(y_1, y_2, \ldots, y_n)$... Потрібно згладити отриману статистичну залежність апроксимуючою кривою $\overline{y}_x = \varphi(x)$. Будемо вважати, що відхилення статистичних даних від апроксимуючої кривої $y_i - \varphi(x_i)$ (рис. 9.3) обумовлені помилками вимірювання, і виходить, розподілені нормально. Тоді результативна ознака Y при кожному $X=x_i$ є випадковою величиною Y_i, розподіленою нормально з параметрами $\varphi(x_i)$ і σ. Параметр розподілу σ характеризує точність виміру Y в i-му досліді. Будемо вважати, що вимірювання у всіх дослідах проводилися з однаковою точністю, тоді σ для всіх Y_1, Y_2, …, Y_n та сама. Закон розподілу Y_i запишемо в такий спосіб:

$$f(Y_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{[y_i - \varphi(x_i)]^2}{2\sigma^2}\right\}.$$

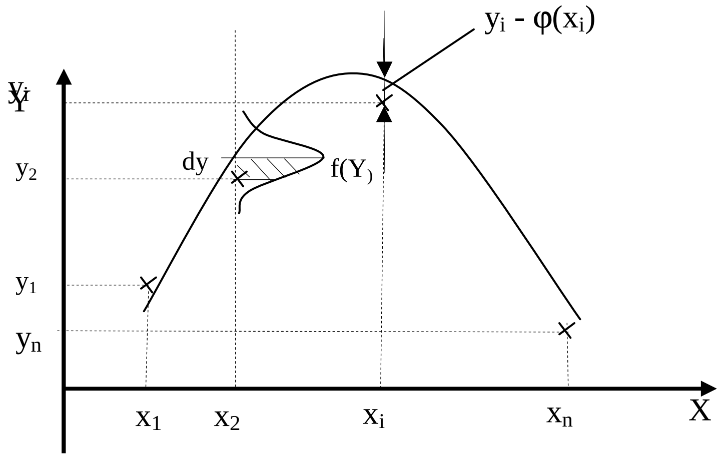


Рис. 9.3 – Розподіл відхилень статистичних даних від апроксимуючої кривої y_i - $\varphi(x_i)$

Імовірність того, що Y_i потрапила в інтервал dy дорівнює елементу імовірності (рис. 9.3):

$$P\{y_i - dy < Y_i < y_i + dy\} = f(Y_i)*dy.$$

Імовірність того, що результативна ознака Y прийняла значення y_1, y_2, …, y_n визначимо за теоремою множення:

$$P\{\, y_1 = y_1, y_2 = y_2, \ldots, y_n = y_n \,\} = \prod_{i=1}^{n} f(Y_i)dy = \left(\frac{dy}{\sigma\sqrt{2\pi}}\right)^n exp\left\{-\sum_{i=1}^{n}\frac{[\,y_i - \varphi(x_i)\,]^2}{2\sigma^2}\right\}.$$

Ця імовірність буде найбільшою, коли аргумент експоненти $\left\{-\sum_{i=1}^{n}\frac{[\,y_i - \varphi(x_i)\,]^2}{2\sigma^2}\right\}$ прийме найменше значення. Таким чином, при використанні МНК вимога найкращого узгодження апроксимуючої кривої $\overline{y}_x = \varphi(x)$ з дослідними даними зводиться до того, щоб сума квадратів відхилень цієї кривої від експериментальних точок оберталася в мінімум:

$$\sum_{i=1}^{n}(\,y_i - y_{ip}\,)^2 \to min\,. \qquad (9.4)$$

де y_i – значення результативної ознаки Y, отримані в результаті спостережень; y_{ip} - розрахункові значення результативної ознаки Y, отримані на підставі аналітичного вираження кривої, що згладжує $\overline{y}_x = \varphi(x)$.

Якщо всі вимірювання проводилися з однаковою точністю і помилки вимірювань розподілені за нормальним законом, то знайдена залежність буде найбільш імовірною з усіх можливих в даному класі функцій.

З огляду на те, що $y_{ip} = \varphi(x_i)$, вираз (9.4) можна записати у вигляді:

$$\sum_{i=1}^{n} [\, y_i - \varphi(\, x_i\,)]^2 \to min \, . \tag{9.5}$$

Невідомі параметри шуканої залежності визначають, записавши її не тільки як функцію аргументу x, але і як функцію невідомих параметрів a_j, $j = \overline{1,m}$:

$$\sum_{i=1}^{n} [\, y_i - \varphi(\, x_i\, , a_1\, , a_2\, ,..., a_j\, ,..., a_m\,)]^2 \to min\, , \tag{9.6}$$

де m - число шуканих параметрів.

Візьмемо часткові похідні від виразу (9.6) за параметрами a_j і, дорівнявши їх нулю, дістанемо систему m+1 нормальних рівнянь з m+1 невідомими, розв'язання якої дає шукані параметри a_j, що задовольняють умові (9.5):

$$-2\sum_{i=1}^{n}\left[y_i - \varphi(\, x_i\, , a_0\, , a_1\, ,\ldots, a_j\, ,\ldots, a_m\,)\right]\frac{\partial\varphi}{\partial a_j} = 0\, , \quad j = \overline{1,m}\, .$$

Розв'язання отриманої системи нормальних рівнянь залежить від конкретного вигляду залежності $\overline{y}_x = \varphi(x)$.

Отримаємо для лінійного рівняння регресії (9.3) методом найменших квадратів вираз для коефіцієнта регресії ρ_{yx} і вільного члена b. Для цього підставимо в (9.6) вираз (9.3) і дістанемо:

$$\sum_{i=1}^{n} [\, y_i - \rho_{yx} x_i - b\,]^2 \to min\, .$$

Для відшукання мінімуму візьмемо похідні за параметрами ρ_{yx} і b і, дорівнявши їх до нуля, одержимо систему нормальних рівнянь:

$$\left.\begin{aligned} 2\sum_{i=1}^{n} [\, y_i - \rho_{yx} x_i - b\,] * x_i &= 0 \\ 2\sum_{i=1}^{n} [\, y_i - \rho_{yx} x_i - b\,] &= 0 \end{aligned}\right\}, \tag{9.7}$$

з якої в результаті перетворень маємо:

$$\left.\begin{aligned} \rho_{yx} * \sum x_i^2 + b * \sum x_i &= \sum x_i y_i \\ \rho_{yx} * \sum x_i + nb &= \sum y_i \end{aligned}\right\}. \tag{9.8}$$

Виразимо ρ_{yx} і b

$$\rho_{yx} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2}; \qquad (9.9)$$

$$b = \frac{\sum x_i^2 * \sum y_i - \sum x_i * \sum x_i y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2}. \qquad (9.10)$$

Приклад 9.1. Нехай у результаті досліду отримані такі експериментальні дані:

x_i	1	2	3
y_i	1	3	4

Потрібно визначити параметри лінійної і квадратичної залежностей для X і Y.

Розв'язання. Нанесемо на координатну площину точки з координатами (x_i, y_i) (рис. 9.4).

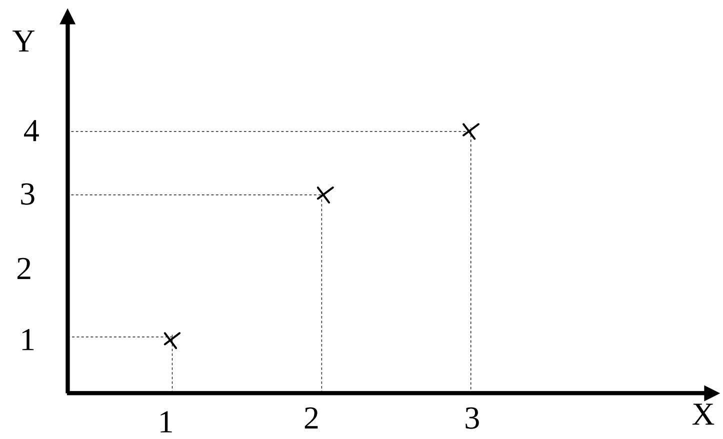


Рис. 9.4 - Побудова поля кореляції

Із графіка видно, що точки не лежать на одній прямій і що із зростанням X У має тенденцію до зростання.

а) нехай шукана залежність - лінійна:

$$y = a_0 + a_1 x,$$

підставивши її в (9.5), дістанемо:

$$\sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2 \to min.$$

Візьмемо часткові похідні за параметрами a_1 і a_0 і дорівняємо їх до нуля:

$$\begin{cases} 2*\Sigma(y_i - a_0 - a_1 x_i)*(-1) = 0; \\ 2*\Sigma(y_i - a_0 - a_1 x_i)*(-x_i) = 0; \end{cases}$$

виконаємо перетворення:

$$\begin{cases} \Sigma a_0 + \Sigma a_1 x_i = \Sigma y_i \\ \Sigma a_0 x_i + \Sigma a_1 x_i^2 = \Sigma y_i x_i \end{cases}$$

$$\begin{cases} n a_0 + a_1 \Sigma x_i = \Sigma y_i \\ a_0 \Sigma x_i + a_1 \Sigma x_i^2 = \Sigma y_i x_i \end{cases}$$

Підставимо значення: n=3; Σx_i =6; Σx_i^2 =14; Σy_i = 8; $\Sigma y_i x_i$ = 19 і визначимо параметри.

$$\begin{cases} a_0 + 6 a_1 = 8 \\ 6a_0 + 14a_1 = 19. \end{cases}$$

$$\begin{cases} a_0 = -1/3 \\ a_1 = 1{,}5. \end{cases}$$

Шукана залежність має вигляд:

$$y = -1/3 + 1{,}5\ x.$$

Отримана лінійна залежність є найбільш імовірною з лінійних залежностей.

Протабулюємо її і побудуємо графік

x_i	1	2	3
y_i	1	3	4
$y_i^т$	1,17	2,67	4,17

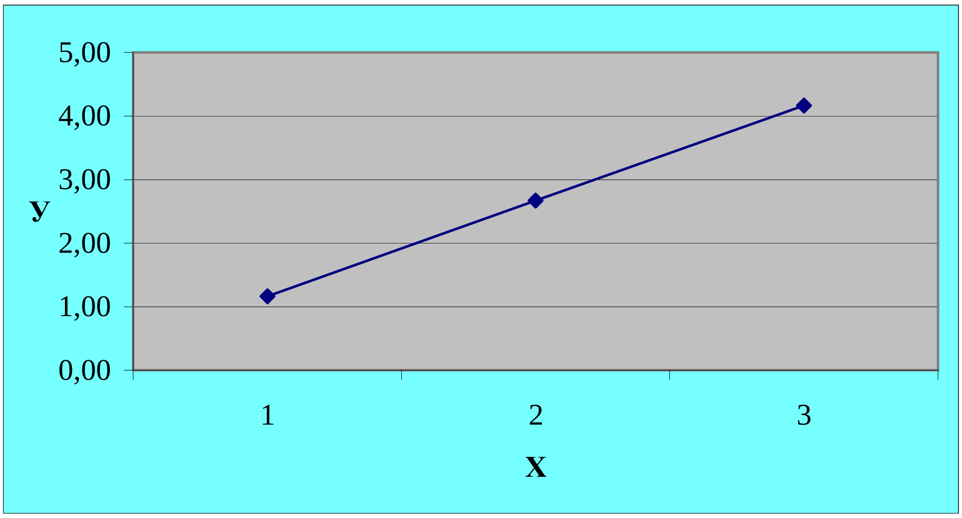


Рис. 9.5 - Лінійна залежність y = -1/3 + 1,5 x

б) нехай шукана залежність – квадратична: $Y = a_0 + a_1x + a_2x^2$, підставимо її в (9.6):

$$\sum_{i=1}^{n} (y_i - a_0 - a_1 x_i - a_2 x_i^2)^2 \rightarrow min ,$$

візьмемо часткові похідні й дорівняємо їх до нуля:

$$\begin{cases} 2*\Sigma(y_i - a_0 - a_1x_i - a_2x_i^2)*(-1) = 0 \\ 2*\Sigma(y_i - a_0 - a_1x_i - a_2x_i^2)*(-x_i) = 0 \\ 2*\Sigma(y_i - a_0 - a_1x_i - a_2x_i^2)*(-x_i^2) = 0, \end{cases}$$

виконаємо перетворення:

$$\begin{cases} \Sigma y_i - \Sigma a_0 - \Sigma a_1x_i - \Sigma a_2x_i^2 = 0 \\ \Sigma y_i\, x_i - \Sigma\, a_0\, x_i - \Sigma a_1x_i^2 - \Sigma a_2x_i^3 = 0 \\ \Sigma y_i\, x_i^2 - \Sigma a_0\, x_i^2 - \Sigma a_1x_i^3 - \Sigma a_2x_i^4 = 0, \end{cases}$$

$$\begin{cases} na_0 + a_1\Sigma x_i + a_2\Sigma x_i^2 = \Sigma y_i \\ a_0\,\Sigma x_i + a_1\Sigma x_i^2 + a_2\Sigma x_i^3 = \Sigma y_i\, x_i \\ a_0\,\Sigma x_i^2 + a_1\Sigma x_i^3 + a_2\Sigma x_i^4 = \Sigma y_i\, x_i^2. \end{cases}$$

Підставимо значення: $n=3$; $\Sigma x_i =6$; $\Sigma x_i^2 =14$; $\Sigma y_i = 8$; $\Sigma y_i\, x_i = 19$;$\Sigma x_i^3 =36$; $\Sigma x_i^4 =98$; $\Sigma y_i\, x_i^2 = 49$ і визначимо параметри

$$\begin{cases} 3a_0 +6\, a_1 + 14a_2 = 8 \\ 6a_0 + 14a_1 + 36a_2 =19 \\ 14a_0 + 36a_1 + 98a_2 = 49. \end{cases}$$

Звідки : a_2=-0,091; a_1=1,273; a_0=0,0455.

Таким чином, найбільш імовірна квадратична залежність матиме вигляд

$$y = 0{,}0455 + 1{,}273x- 0{,}091x^2.$$

Протабулюємо її, збільшивши для наочності число точок:

x_i	0	1	2	3	4	5
$y_i^т$	0,05	1,23	2,23	3,05	3,68	4,14

Отримаємо графік

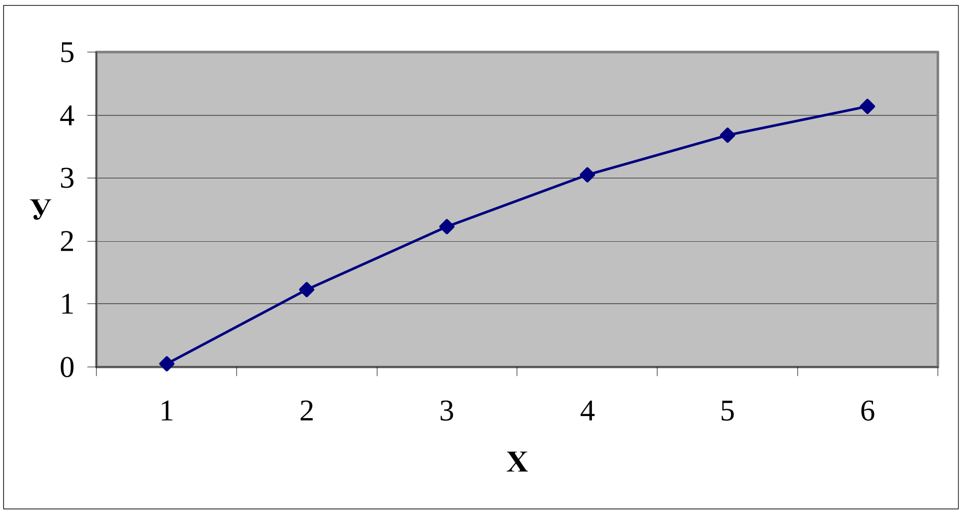


Рис. 9.6 - Квадратична залежність $y = 0{,}0455 + 1{,}273x- 0{,}091x^2$

Приклад 9.2. Визначити параметри лінійної регресії для системи випадкових величин X і Y, результати вимірювання яких наведені в прикладі 8.7.

№ досліду	x_i	y_i
1	4	0,041
2	8	0,05
3	10	0,081
4	14	0,104
5	16	0,12
6	20	0,139
7	19	0,154
8	23	0,18
9	26	0,208
10	30	0,241
11	31	0,25
12	36	0,269
13	37	0,301

Розв'язання. Побудуємо поле кореляції (рис. 9.7). Очевидно, що статистичні дані добре згладжуються лінійною залежністю.

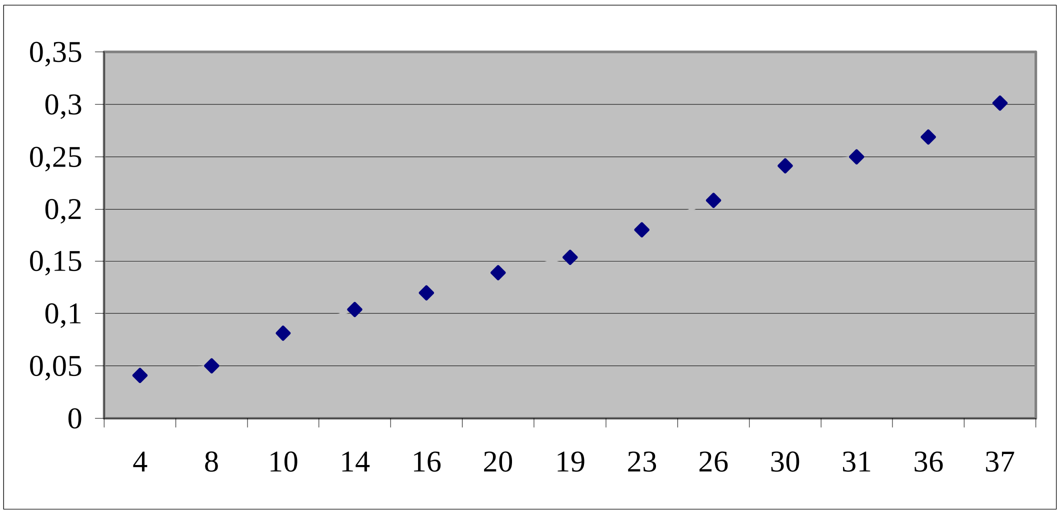


Рис. 9.7 - Побудова поля кореляції

Для визначення параметрів лінійної залежності між X і Y скористаємося формулами (9.9) і (9.10). Проміжні обчислення зробимо в таблиці

№ досліду	x_i	y_i	x_i^2	x_iy_i
1	4	0,041	16	0,164
2	8	0,05	64	0,4
3	10	0,081	100	0,81
4	14	0,104	196	1,456
5	16	0,12	256	1,92
6	20	0,139	400	2,78
7	19	0,154	361	2,926
8	23	0,18	529	4,14
9	26	0,208	676	5,408
10	30	0,241	900	7,23
11	31	0,25	961	7,75
12	36	0,269	1296	9,684
13	37	0,301	1369	11,137
Сума	Σx_i=274	Σy_i=2,138	Σx_i^2= 7124	Σx_i*y_i= 55,805

Дістанемо

$$\rho_{yx} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} = \frac{13*55{,}08 - 274*2{,}138}{13*7124 - 274^2} = 0{,}00796\,;$$

$$b = \frac{\sum x_i^2 * \sum y_i - \sum x_i * \sum x_i y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2} = \frac{7124*2{,}138 - 274*55{,}8}{13*7124 - 274^2} = -0{,}0034\,.$$

Таким чином, шукана залежність має вигляд

$$y = 0{,}00796x - 0{,}0034.$$

Для визначення параметрів лінійної залежності між X і Y можна скористатися функціями Microsoft Excel. Визначимо коефіцієнти ρ_{yx} і b за допомогою функції ЛИНЕЙН. Функція ЛИНЕЙН розраховує статистику для ряду із застосуванням методу найменших квадратів, щоб обчислити пряму лінію, яка щонайкраще апроксимує наявні дані. Функція повертає масив, який описує отриману пряму. Оскільки повертається масив значень, функція повинна задаватися у вигляді формули масиву. Синтаксис функції:

=ЛИНЕЙН(известные_значения_у;известные_значения_х;конст;статистика),

де Известные_значення_у — множина значень у; Известные_значення_х — необов'язкова множина значень х, якщо Известные_значення_х опущені, то передбачається, що це масив {1;2;3;...} такого ж розміру, як і Известные_значення_у; Конст — логічне значення, що вказує, чи потрібно, щоб константа b дорівнювала 0. Якщо Конст має значення ИСТИНА або опущена, то b обчислюється звичайним способом. Якщо аргумент Конст має значення ЛОЖЬ, то b вважається рівним 0; Статистика - логічне значення, що вказує, чи потрібно повернути додаткову статистику по регресії.

Оскільки функція повертає масив, виділимо в аркуші Excel дві комірки, в яких розмістяться параметри ρ_{yx} і b (рис. 9.8). Як аргумент Известные_значення_y вкажемо діапазон C2:C14, як аргумент Известные_значення_x - діапазон B2:B14. Щоб отримати ненульове значення b, в аргумент Конст введемо слово ИСТИНА. Оскільки нам не потрібні інші дані, в аргумент Статистика введемо слово ЛОЖЬ.

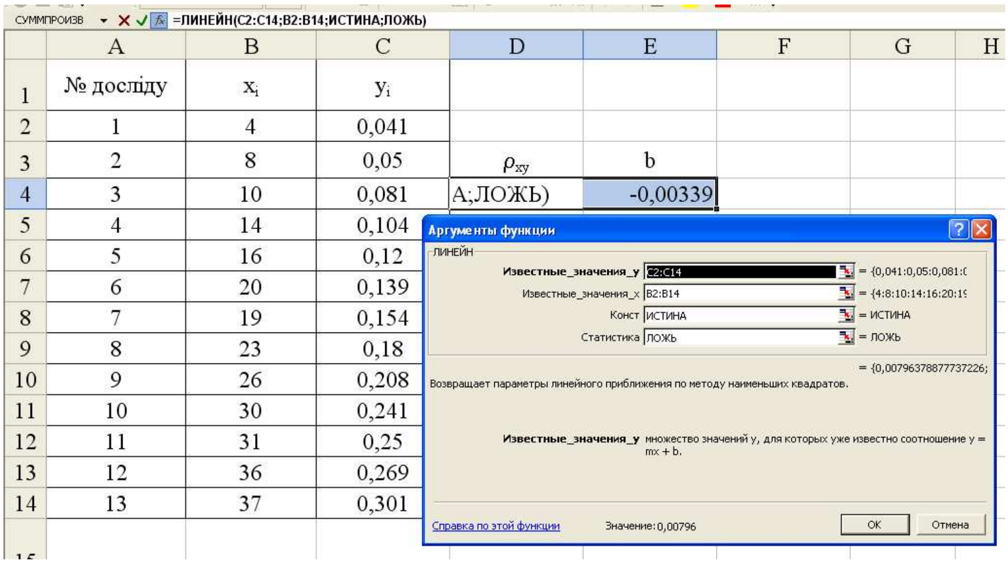


Рис. 9.8 - Введення аргументів функції ЛИНЕЙН

Для завершення введення функції скористаємось командою масиву: натискання функціональної клавіші F2, а потім комбінації клавіш Ctrl+Shift+Enter. Результат показаний на рис. 9.9.

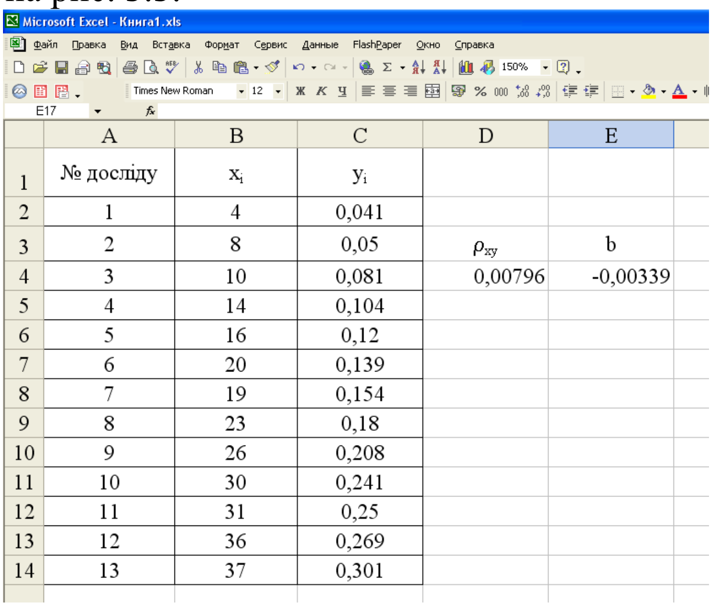


Рис. 9.9 - Визначення параметрів ρ_{yx} і b з використанням функції ЛИНЕЙН

Побудуємо графік отриманої залежності y = 0,00796x - 0,0034 (рис. 9.10).

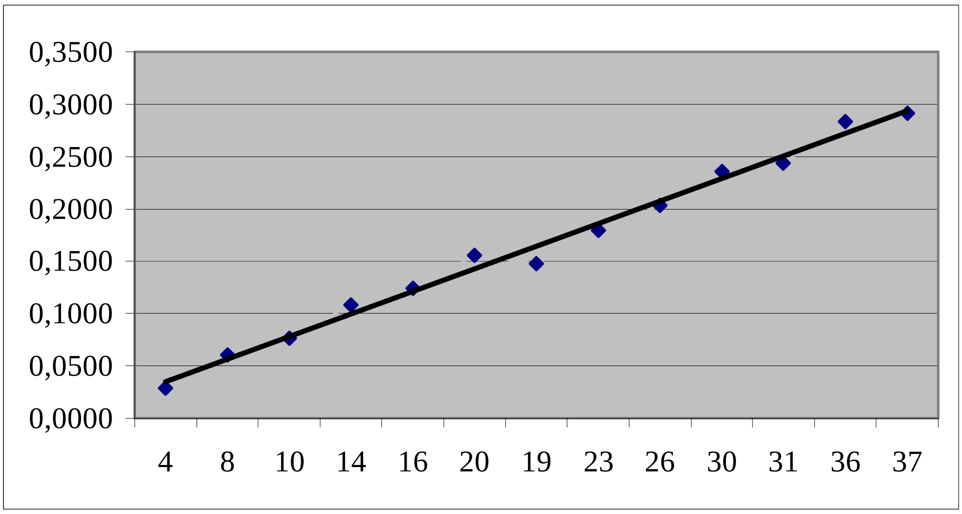


Рис. 9.10 - Графік залежності y = 0,00796x - 0,0034

Приклад 9.3. Визначити параметри лінійної регресії для системи випадкових величин I і U, результати вимірювання яких наведені в прикладі 8.9. Для обчислень скористатися функціями Microsoft Excel.

Номер досліду	Ii	Ui	Номер досліду	Ii	Ui
1	-18	-30	11	4	-8
2	-16	-32	12	6	22
3	-14	-10	13	8	-2
4	-12	-8	14	10	10
5	-10	-15	15	12	3
6	-8	10	16	14	18
7	-6	5	17	16	55
8	-4	2	18	18	28
9	-2	3	19	20	22
10	2	8	20	22	62

Розв'язання. Для визначення параметрів лінійної регресії скористаємося статистичними функціями Microsoft Excel НАКЛОН, ОТРЕЗОК і ТЕНДЕНЦИЯ.

Функція НАКЛОН дає значення параметра ρ_{yx}. Вона повертає нахил лінії лінійної регресії для точок даних. Нахил визначається як частка від ділення відстані по вертикалі на відстань по горизонталі між двома будь-якими точками прямої, тобто нахил - це швидкість зміни значень уздовж прямої (рис. 9.11). Синтаксис:

НАКЛОН (известные_значения_y;известные_значения_x).

Аргументи повинні бути числами або іменами, масивами або посиланнями, що містять числа. Якщо аргумент, що є масивом або посиланням, містить текст, логічні значення або порожні комірки, ці значення ігноруються; комірки, які містять нульові значення, враховуються. Якщо известные_значения_y і известные_значения_x порожні або містять різне число точок даних, то функція НАКЛОН повертає значення помилки #Н/Д.

	A	B	C	D	E	F	G
1	Номер досліду	Ii	Ui				
2	1	-18	-30				
3	2	-16	-32				
4	3	-14	-10				
5	4	-12	-8				
6	5	-10	-15				
7	6	-8	10			ρ_{yu}	=НАКЛОН(C2:C21;B2:B21)
8	7	-6	5				
9	8	-4	2				
10	9	-2	3				
11	10	2	8				
12	11	4	-8				
13	12	6	22				
14	13	8	-2				
15	14	10	10				
16	15	12	3				
17	16	14	18				
18	17	16	55				
19	18	18	28				
20	19	20	22				
21	20	22	62				

Рис. 9.11 - Використання функції НАКЛОН

Дістанемо ρ_{yx}.= 1,525.

Значення параметра b визначимо за допомогою функції ОТРЕЗОК (рис. 9.12). Функція ОТРЕЗОК обчислює точку перетинання лінії з віссю у. Точка перетинання знаходиться на оптимальній лінії регресії, проведеній через известные_значения_x і известные_значения_y. Функція ОТРЕЗОК використовується, коли потрібно визначити значення залежної змінної при значенні незалежної змінної, рівному нулю. Синтаксис:

ОТРЕЗОК (известные_значения_x; известные_значения_y).

Аргументи повинні бути числами або іменами, масивами або посиланнями, що містять числа. Якщо аргумент, що є масивом або посиланням, містить текст, логічні значення або порожні комірки, ці значення ігноруються; комірки, що містять нульові значення, враховуються. Якщо известные_значения_y і известные_значения_x містять різну кількість точок даних або зовсім не містять точок даних, то функція ОТРЕЗОК повертає значення помилки #Н/Д.

G9 fx =ОТРЕЗОК(C2:C21;B2:B21)

	A	B	C	D	E	F	G
1	Номер досліду	Ii	Ui				
2	1	-18	-30				
3	2	-16	-32				
4	3	-14	-10				
5	4	-12	-8				
6	5	-10	-15				
7	6	-8	10			ρ_{yu}	1,525
8	7	-6	5				
9	8	-4	2			b	3,948
10	9	-2	3				
11	10	2	8				
12	11	4	-8				
13	12	6	22				
14	13	8	-2				
15	14	10	10				
16	15	12	3				
17	16	14	18				
18	17	16	55				
19	18	18	28				
20	19	20	22				
21	20	22	62				

Рис. 9.12 - Використання функції ОТРЕЗОК

Дістанемо b = 3,948. Таким чином, шукана залежність має вигляд

$$u = 1{,}525i + 3{,}948.$$

Використаємо функцію ТЕНДЕНЦИЯ для одержання значень u, обчислених на підставі апроксимуючої залежності (рис. 9.13 і 9.14). Функція ТЕНДЕНЦИЯ повертає значення відповідно до лінійного тренда. Апроксимує прямою лінією (за методом найменших квадратів) масиви даних і повертає значення у, відповідно до цієї прямої для заданого масиву. Синтаксис:

ТЕНДЕНЦИЯ(изестные_значения_y;известные_значения_x;новые_значения_x;конст),

де Новые_значения_x — значення x, для яких ТЕНДЕНЦИЯ повертає відповідні значення y. Новые_значения_x повинні містити стовпець (або рядок) для кожної незалежної змінної, як і известные_значения_x. Якщо новые_значения_x опущені, то передбачається, що вони збігаються з известные_значения_x. Якщо опущені обидва масиви известные_значения_x і новые_значения_x, то передбачається, що це масив {1;2;3;...} такого ж розміру, що і известные_значения_y; Конст — логічне значення, що вказує, чи потрібно, щоб константа b дорівнювала 0. Якщо Конст має значення ИСТИНА або опущена, то b обчислюється звичайним способом.

D2 fx {=ТЕНДЕНЦИЯ(C2:C21;B2:B21)}

	A	B	C	D	E	F	G
1	Номер досліду	Ii	Ui				
2	1	-18	-30	-23,4967			
3	2	-16	-32	-20,4472			
4	3	-14	-10	-17,3978			
5	4	-12	-8	-14,3484			
6	5	-10	-15	-11,299			
7	6	-8	10	-8,24956		ρ_{yu}	1,525
8	7	-6	5	-5,20014			
9	8	-4	2	-2,15073		b	3,948
10	9	-2	3	0,898693			
11	10	2	8	6,997529			
12	11	4	-8	10,04695			
13	12	6	22	13,09637			
14	13	8	-2	16,14578			
15	14	10	10	19,1952			
16	15	12	3	22,24462			
17	16	14	18	25,29404			
18	17	16	55	28,34346			
19	18	18	28	31,39287			
20	19	20	22	34,44229			
21	20	22	62	37,49171			

Рис. 9.13. Використання функції ТЕНДЕНЦИЯ

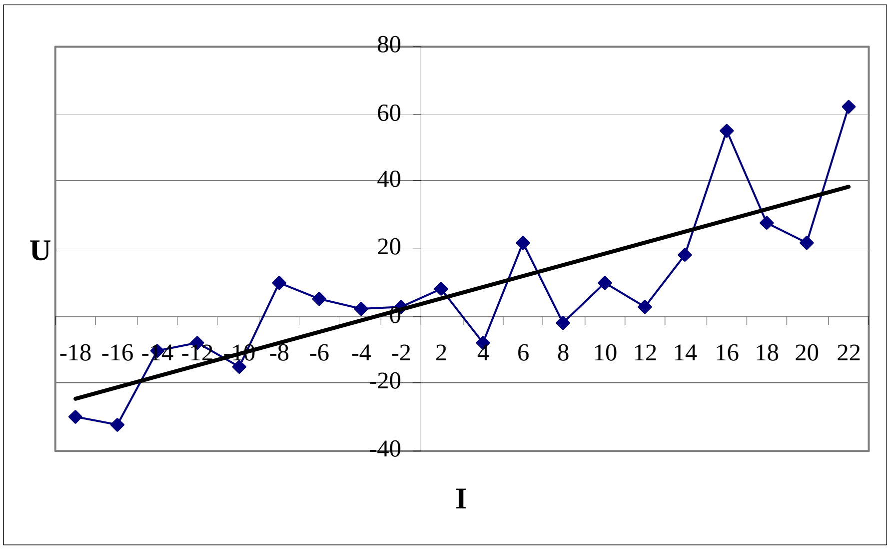


Рис. 9.14 - Графік залежності u = 1,525i + 3,948

Із графіка на рис. 9.14 видно, що розкид статистичних значень U від прямої, що згладжує, досить великий, однак з усіх гладких залежностей лінійна залежність найбільш імовірна.

Приклад 9.4. Визначити параметри лінійної і квадратичної регресії для системи випадкових величин U і Z, використовуючи функції Microsoft Excel.

Номер досліду	U_i	Z_i	Номер досліду	U_i	Z_i
1	1	2	11	8	12
2	1	3	12	8	27,2
3	1	8	13	10	52,7
4	2	14,4	14	10	39
5	2	5	15	11	90,1
6	2	24	16	12	160
7	4	11,3	17	12	130,3
8	4	17,2	18	14	159
9	6	10	19	16	190
10	6	12	20	20	195

Розв'язання. З таблиці видно, що в статистичних даних наведені результати паралельних дослідів. Усереднивши їх і визначивши дисперсії, отримаємо:

Номер групи	U_i	Групові середні Z_i	Групові дисперсії Z_i
1	1	4,33	10,33
2	5	14,47	90,25
3	9	14,25	17,41
4	13	11,00	2,00
5	17	19,60	115,52
6	21	45,85	93,85
7	25	102,55	310,01
8	29	145,15	441,04
9	33	153,50	60,50
10	37	183,00	98,00
11	41	196,50	4,50

Визначимо параметри лінійної регресії. Скористаємося для цього статистичними функціями Microsoft Excel НАКЛОН і ОТРЕЗОК (рис. 9.15).

27							
28	Номер групи	Ui	Групові середні Zi	Групові дисперсії			
29	1	1	4,33	10,33	ρ_{uz}	5,4636	0,940647
30	2	5	14,47	90,25			
31	3	9	14,25	17,41	b	-33,808	
32	4	13	11,00	2,00			
33	5	17	19,60	115,52			
34	6	21	45,85	93,85			
35	7	25	102,55	310,01			
36	8	29	145,15	441,04			
37	9	33	153,50	60,50			
38	10	37	183,00	98,00			
39	11	41	196,50	4,50			

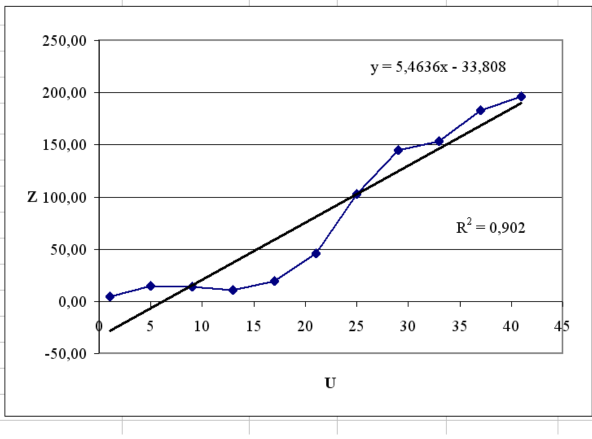


Рис. 9.15 - Побудова лінійної регресії і визначення параметрів

Таким чином, найбільш імовірна лінійна залежність має вигляд

$$z = 5{,}46u - 33{,}8.$$

Виділивши діапазон комірок M29:N39, побудуємо точковий графік. Потім клацнувши правою кнопкою миші на графіку, виберемо команду «Додати лінію тренда». У вікні «Лінія тренда», що з'явилося, на вкладці «Тип» виберемо «Лінійна» , на вкладці «Параметри» включимо прапорці «показувати рівняння на діаграмі» і «помістити на діаграму величину вірогідності апроксимації R^2».

Тепер визначимо параметри квадратичної регресії.

$$z = a_0 + a_1 u + a_2 u^2$$

Їх можна знайти із системи рівнянь, отриманої в прикладі 9.1:

$$\begin{cases} n a_0 + a_1 \Sigma x_i + a_2 \Sigma x_i^2 = \Sigma y_i \\ a_0\, \Sigma x_i + a_1 \Sigma x_i^2 + a_2 \Sigma x_i^3 = \Sigma y_i\, x_i \\ a_0\, \Sigma x_i^2 + a_1 \Sigma x_i^3 + a_2 \Sigma x_i^4 = \Sigma y_i\, x_i^2. \end{cases}$$

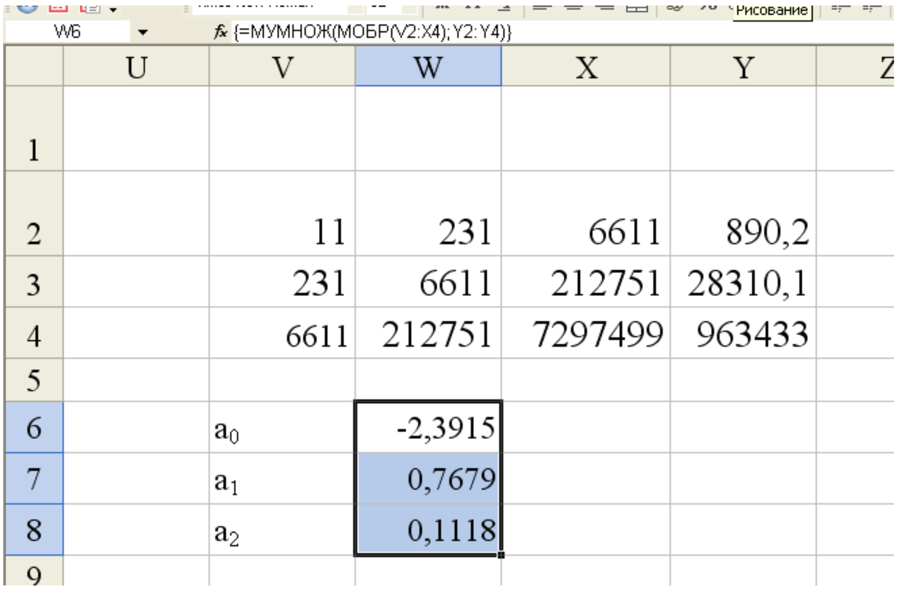
Зробимо проміжні розрахунки в таблиці

Номер групи	**U_i**	**Групові середні Z_i**	**u_i^2**	**u_i^3**	**u_i^4**	**uz**	**u^2z**
1	1	4,33	1	1	1	4,33	4,33
2	5	14,47	25	125	625	72,33	361,67
3	9	14,25	81	729	6561	128,25	1154,25
4	13	11,00	169	2197	28561	143,00	1859,00
5	17	19,60	289	4913	83521	333,20	5664,40
6	21	45,85	441	9261	194481	962,85	20219,85
7	25	102,55	625	15625	390625	2563,75	64093,75
8	29	145,15	841	24389	707281	4209,35	122071,15
9	33	153,50	1089	35937	1185921	5065,50	167161,50
10	37	183,00	1369	50653	1874161	6771,00	250527,00
11	41	196,50	1681	68921	2825761	8056,50	330316,50
Сума	231	890,2	6611	212751	7297499	28310,0667	963433,4

У результаті дістанемо систему рівнянь

$$\begin{cases} a_0 11 + a_1 231 + a_2 6611 = 890,2 \\ a_0\, 231 + a_1 6611 + a_2 212751 = 28310,0667 \\ a_0\, 6611 + a_1 212751 + a_2 7297499 = 963433,4. \end{cases}$$

Вирішимо систему, скориставшись формулою для обчислення коренів =МУМНОЖ(МОБР(V2:X4);Y2:Y4), натиснувши потім комбінацію клавіш ctrl+shift+enter (рис. 9.16)


W6 {=МУМНОЖ(МОБР(V2:X4);Y2:Y4)}

	U	V	W	X	Y	Z
1						
2		11	231	6611	890,2	
3		231	6611	212751	28310,1	
4		6611	212751	7297499	963433	
5						
6		a_0	-2,3915			
7		a_1	0,7679			
8		a_2	0,1118			
9						

Рис. 9.16. Розв’язання системи рівнянь

Таким чином, найбільш імовірна квадратична залежність має вигляд

$$z = -2,39 + 0,7679u + 0,1118u^2.$$

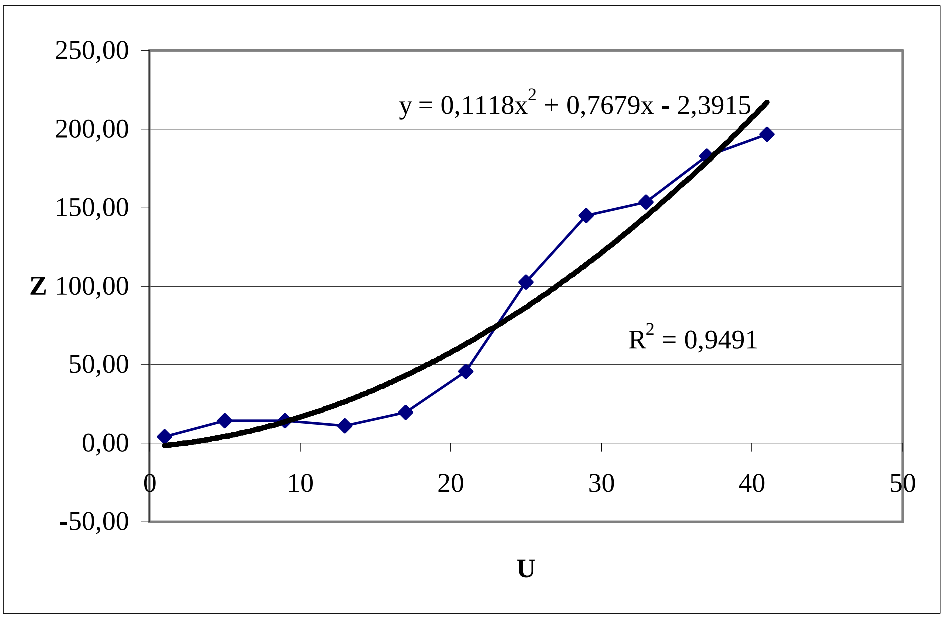


Рис. 9.17 - Побудова квадратичної регресії і визначення параметрів

Виділивши діапазон комірок M29:N39, побудуємо точковий графік і додамо лінію тренда, вибравши тип «Поліноміальна» з показником степені, рівним 2. Помістимо на діаграмі рівняння і величину R^2 (рис. 9.17).

Вибірковий коефіцієнт кореляції

Для оцінки тісноти лінійної кореляційної залежності слугує вибірковий коефіцієнт кореляції. Для його визначення підставимо у вираз (9.8), використане для одержання параметрів лінійної залежності за методом найменших квадратів, наступні співвідношення:

$$\overline{x} = \frac{\sum x_i}{n}\text{, звідки } \sum x_i = \overline{x}n;$$

$$\overline{y} = \frac{\sum y_i}{n}\text{, звідки } \sum y_i = \overline{y}n;$$

$$\overline{x^2} = \frac{\sum x_i^2}{n}\text{, звідки } \sum x_i^2 = \overline{x^2}n;$$

Дістанемо

$$\left.\begin{aligned} &\rho_{yx} * \overline{x^2}n + b * \overline{x}n = \sum x_i y_i \\ &\rho_{yx} * \overline{x} + b = \overline{y} \end{aligned}\right\}, \qquad (9.10)$$

з другого рівняння виразимо b:

$$b = \overline{y} - \rho_{yx} * \overline{x} \qquad (9.11)$$

і, підставивши його в перше рівняння, знайдемо коефіцієнт регресії:

$$\rho * \sum x_i^2 + (\overline{y} - \rho * \overline{x})\sum x_i = \sum x_i y_i$$

$$\rho * \left(\sum x_i^2 - \overline{x} * \sum x_i\right) = \sum x_i y_i - \overline{y} * \sum x_i$$

$$\rho = \frac{\sum x_i y_i - \overline{y} * \sum x_i}{\sum x_i^2 - \overline{x} * \sum x_i} = \frac{\sum x_i y_i - n\overline{y}\overline{x}}{n\overline{x^2} - n\overline{x}^2} = \frac{\sum x_i y_i - n\overline{y}\overline{x}}{n * \sigma_x^2}. \qquad (9.12)$$

Помножимо рівність (9.12) на дріб $\frac{\sigma_x}{\sigma_y}$, тоді

$$\rho * \frac{\sigma_x}{\sigma_y} = \frac{\sum x_i y_i - n\overline{y}\overline{x}}{n * \sigma_x \sigma_y}, \qquad (9.13)$$

де $\rho * \frac{\sigma_x}{\sigma_y} = r_в$ - вибірковий коефіцієнт кореляції.

Підставивши в рівняння лінійної регресії $\overline{y}_x = \rho_{yx} x + b$ вираз для b (9.11), отримаємо його в наступному вигляді:

$$\overline{y_x}-\overline{y}=\rho_{xy}*(x-\overline{x}) \quad \text{або} \quad \overline{y_x}-\overline{y}=r_в\frac{\sigma_y}{\sigma_x}*(x-\overline{x}). \qquad (9.14)$$

Коефіцієнт кореляції $r_в$ має важливе значення. Він дозволяє оцінити величину лінійного зв'язку між двома випадковими величинами X і Y. Покладемо в рівнянні (9.13) $r_в = 0$, тоді

$$\overline{y_x}-\overline{y}=0,$$

або

$$\overline{y_x}=\overline{y},$$

тобто при $r_в = 0$ всі умовні середні дорівнюють вибірковій середній, а значить при зміні фактора X результативна ознака Y не змінюється, і графік рівняння регресії паралельний осі абсцис. Це говорить про те, що Y не залежить від X, між ними немає лінійного зв'язку. Однак X і Y можуть бути зв'язані нелінійним зв'язком, що може виявитися як кореляційним, так і функціональним.

Дисперсія результативної ознаки Y в точці $X=x_i$ відносно свого умовного середнього S_y визначається за формулою

$$S_y=D_y(1-r_в^2), \qquad (9.15)$$

де D_y - дисперсія Y відносно загальної середньої.

Покладемо в цій формулі $r_в = 1$, тоді

$$S_y=0,$$

тобто при $r_в = 1$ розсіювання значень результативної ознаки Y в кожній точці відсутнє, рівняння (9.14) матиме вигляд $\overline{y_x}-\overline{y}-\frac{\sigma_y}{\sigma_x}*(x-\overline{x})=0$, тобто будь-яка пара чисел x і y йому задовольняє. Звідси випливає, що при $r_в = 1$ між X і Y існує функціональний лінійний зв'язок.

З формули (9.14) також виходить, що зі збільшенням $r_в$ дисперсія результативної ознаки Y відносно умовної середньої S_y зменшується, тобто зменшується розсіювання навколо умовних середніх, а значить тіснота зв'язку збільшується.

Таким чином, вибірковий коефіцієнт кореляції приймає значення від -1 до +1 і характеризує тісноту лінійного зв'язку між ознаками у вибірці. Якщо $r_в = 0$, то лінійний зв'язок відсутній, чим ближче значення $|r_в|$ до одиниці, тим тісніше зв'язок, при $|r_в| = 1$ він стає функціональним.

Вибіркове кореляційне відношення

Для оцінки тісноти нелінійного кореляційного зв'язку застосовують вибіркове кореляційне відношення η. Вибірковим **кореляційним відношенням** Y до X називається відношення міжгрупового середнього квадратичного відхилення до загального середнього квадратичного відхилення результативної ознаки Y:

$$\eta_{yx}=\frac{\sigma_{\overline{y_x}}}{\sigma_y} \qquad (9.16)$$

де $\sigma_{\overline{y_x}}$ - міжгрупове середнє квадратичне відхилення. Воно визначається як квадратний корінь з міжгрупової дисперсії за формулою

$$\sigma_{\overline{y_x}} = \sqrt{D_{межгр}} = \sqrt{\frac{\sum N_j(\overline{y_j} - \overline{y})^2}{n}}, \qquad (9.17)$$

де $\overline{y_j}$ - умовна середня значень Y j-ї групи; N_j – обсяг j-ї групи.

Міжгрупова дисперсія – це дисперсія групових середніх відносно загальної середньої.

Внутрішньогрупова дисперсія являє собою середнє арифметичне групових дисперсій:

$$D_{вн.гр} = \frac{\sum N_j S_y}{n} \qquad (9.18)$$

Можна показати, що загальна дисперсія результативної ознаки Y є сумою внутрішньогрупової і міжгрупової дисперсій:

$$D_y = D_{вн.гр} + D_{міжгр}. \qquad (9.19)$$

Якщо результативна ознака Y функціонально залежить від фактора X, то кожному певному значенню X відповідає єдине значення Y, і дисперсія відносно умовної середньої $S_y = 0$, відповідно дорівнює нулю середнє арифметичне дисперсій відносно умовної середньої. Таким чином, при функціональному зв'язку між X і Y внутрішньогрупова дисперсія дорівнює нулю, а загальна дисперсія результативної ознаки Y дорівнює міжгруповій дисперсії, отже якщо зв'язок функціональний, то кореляційне відношення дорівнює одиниці:

$$\eta_{yx} = \frac{\sigma_{\overline{y_x}}}{\sigma_y} = 1. \qquad (9.20)$$

Якщо кореляційне відношення дорівнює нулю, $\eta=0$, то між X і Y зв'язок відсутній. Це випливає з того, що в цьому разі міжгрупова дисперсія дорівнює нулю, а значить відсутній розкид умовних середніх відносно загальної середньої. Тобто умовні середні при всіх значеннях X однакові, а значить результативна ознака Y не залежить від фактора X.

Кореляційне відношення має наступні властивості:

його значення лежать у межах від 0 до 1:

$$0 \le \eta \le 1;$$

значення кореляційного відношення більше або дорівнює вибірковому коефіцієнту кореляції:

$$\eta \ge |r_в|;$$

якщо кореляційне відношення дорівнює вибірковому коефіцієнту кореляції, $\eta = |r_в|$, то між X і Y є лінійна кореляційна залежність.

Приклад 9.5. Визначимо значення коефіцієнта кореляції системи випадкових величин X і Y з прикладів 8.7 і 9.2.

Розв'язання. У попередніх прикладах для системи випадкових величин X і Y була визначена лінійна кореляційна залежність. Скористаємося параметрами цієї залежності і визначимо коефіцієнт кореляції за формулою (9.13), розрахувавши попередньо середні квадратичні відхилення σ_x і σ_y (рис. 9.18):

$$\sigma_x = 10{,}6;\ \sigma_y = 0{,}085;$$

$$r_{вxy} = \rho * \frac{\sigma_x}{\sigma_y} = 0{,}00796\frac{10{,}6}{0{,}085} = 0{,}994.$$

F10 fx =E4*E7/F7

	A	B	C	D	E	F
1	№ досліду	x_i	y_i			
2	1	4	0,041	0,0285		
3	2	8	0,05	0,0603	ρ_{xy}	b
4	3	10	0,081	0,0762	0,00796	-0,00339
5	4	14	0,104	0,1081		
6	5	16	0,12	0,1240	σ_x	σ_y
7	6	20	0,139	0,1559	10,602	0,085
8	7	19	0,154	0,1479		
9	8	23	0,18	0,1798		
10	9	26	0,208	0,2037	r_{xy}=	0,994023
11	10	30	0,241	0,2355		
12	11	31	0,25	0,2435		
13	12	36	0,269	0,2833		
14	13	37	0,301	0,2913		

Рис. 9.18 - Розрахунок коефіцієнта кореляції

Значення коефіцієнта кореляції $r_{xy} = 0{,}994$ вказує на те, що між X і Y є тісний лінійний зв'язок. Дійсно, з рис. 9.10 видно, що розкид статистичних значень Y відносно прямої лінії, яка згладжує залежність, незначний.

Приклад 9.6. Визначимо значення коефіцієнта кореляції системи випадкових величин I і U з прикладів 8.9 і 9.3.

Розв'язання. У попередніх прикладах для системи випадкових величин I і U була визначена лінійна кореляційна залежність. Розрахуємо середні квадратичні відхилення σ_i і σ_u і визначимо коефіцієнт кореляції за формулою (9.13), (рис. 9.19):

$$\sigma_i = 12{,}72;\ \sigma_u = 23{,}64;$$

$$r_{вiu} = \rho * \frac{\sigma_i}{\sigma_u} = 1{,}525\frac{12{,}72}{23{,}64} = 0{,}821.$$

G14 =G7*F12/G12

	A	B	C	D	E	F	G
1	Номер досліду	Ii	Ui				
2	1	-18	-30	-23,4967			
3	2	-16	-32	-20,4472			
4	3	-14	-10	-17,3978			
5	4	-12	-8	-14,3484			
6	5	-10	-15	-11,299			
7	6	-8	10	-8,24956		ρ_{yu}	1,525
8	7	-6	5	-5,20014			
9	8	-4	2	-2,15073		b	3,948
10	9	-2	3	0,898693			
11	10	2	8	6,997529		σ_i	σ_u
12	11	4	-8	10,04695		12,72	23,64
13	12	6	22	13,09637			
14	13	8	-2	16,14578		r_{iu}=	0,821
15	14	10	10	19,1952			
16	15	12	3	22,24462			
17	16	14	18	25,29404			
18	17	16	55	28,34346			
19	18	18	28	31,39287			
20	19	20	22	34,44229			
21	20	22	62	37,49171			

Рис. 9.19 - Розрахунок коефіцієнта кореляції для прикладів 8.9 і 9.3

Приклад 9.7. Знайдемо значення коефіцієнта кореляції і кореляційного відношення системи випадкових величин U і Z з прикладу 9.4.

Розв'язання. У прикладі 9.4 для системи випадкових величин U і Z були визначені параметри лінійної кореляційної залежності. Розрахуємо середні квадратичні відхилення σ_u=13,27 і σ_z=76,32 і визначимо коефіцієнт кореляції за формулою (9.13):

$$r_{вuz} = \rho * \frac{\sigma_u}{\sigma_z} = 5{,}464\frac{13{,}27}{77{,}06} = 0{,}94 .$$

Визначимо кореляційне відношення як відношення міжгрупового середнього квадратичного відхилення до загального середнього квадратичного відхилення ознаки Z. Для цього розрахуємо міжгрупове середнє квадратичне відхилення за формулою (9.17):

$$\sigma_{\bar{z}_u} = \sqrt{D_{межгр}} = \sqrt{\frac{\sum N_j(\bar{z}_j - \bar{z})^2}{n}} = \sqrt{5824{,}56} = 76{,}32 ,$$

тоді кореляційне відношення визначиться в такий спосіб:

$$\eta_{uz} = \frac{\sigma_{\bar{z}_u}}{\sigma_z} = \frac{76{,}32}{77{,}06} = 0{,}99 .$$

Відомо, що коли кореляційне відношення дорівнює вибірковому коефіцієнту кореляції, $\eta = |r_в|$, то між фактором і результативною ознакою є лінійна кореляційна залежність. У нашому випадку кореляційне відношення трохи перевищує значення коефіцієнта кореляції, крім того, у прикладі 9.4 при побудові ліній тренда для лінійної і квадратичної залежностей коефіцієнт детермінації, що показує ступінь вірогідності апроксимації, для квадратичної апроксимації дорівнює 0,95, а для лінійної - 0,9. Скоріш за все, залежність між фактором U і результативною ознакою Z нелінійна.

Запитання для самоперевірки:

1. Які задачі вирішують методом кореляційного аналізу?
2. В яких випадках залежність y = f(x) є функціональною, статистичною або кореляційною?
3. Дайте визначення термінів «регресія», «лінія регресії», «рівняння регресії».
4. Поясніть значення термінів «пояснююча змінна», «результативна ознака».
5. З яких міркувань визначають тип кореляційної залежності y= f(x)? Які типи залежностей Ви знаєте?
6. Чим характерна лінійна залежність y = f(x)? Чому її використовують найбільш часто?
7. Як називаються параметри лінійної залежності y = f(x)?
8. Які методи можна використовувати для визначення параметрів рівняння регресії y = f(x)?
9. Якій вимозі задовольняють параметри, визначені за методом найменших квадратів?
10. Назвіть характеристики, що дозволяють оцінити наявність зв'язку між ознакою-фактором і результативною ознакою.
11. Які значення може приймати коефіцієнт кореляції, які висновки можна зробити на підставі цих значень?
12. Які значення може приймати кореляційне відношення, і які висновки можна зробити на підставі цих значень?
13. Що таке кореляційна таблиця?
14. Які параметри визначають за допомогою кореляційної таблиці?

Задачі для самостійного розв'язання

9.1. Результати вимірювань досліджуваної ознаки Y зведені в таблицю

x_i	41	50	81	104	120	139	154	180	208	241	250	269	301
y_i	4	8	10	14	16	20	14	23	26	30	31	36	37

Використовуючи поле кореляції, вибрати клас залежності y=((x), побудувати рівняння регресії Y на X, оцінити тісноту зв'язку між фактором і результативною ознакою.

9.2. Результати вимірювань фактору X і результативної ознаки Y наведені в таблиці

x_i	1	2	3	4	5	6	7	8
y_i	6	17	34	57	86	121	162	209

Користуючись методом найменших квадратів визначити параметри апроксимуючої залежності $y=ax^2+bx+c$.

9.3. У результаті статистичних спостережень зареєстрована залежність u=f(t).

u_i	75	55	40	30	20	15	10	10	5	5
t_i	1	2	3	4	5	6	7	8	9	10

Визначити параметри експонентної апроксимуючої залежності.

Тема 10. ПЕРЕВІРКА СТАТИСТИЧНИХ ГІПОТЕЗ

Статистичні гіпотези

Будь-яка інформація, отримана в результаті обробки статистичних даних, має імовірнісний характер. Зокрема, оцінка генеральної середньої є величиною випадковою, розподіленою нормально з параметрами $\overline{X}$ і $\frac{\sigma}{\sqrt{n}}$. Оцінка генеральної дисперсії також випадкова. Тому будь-який висновок, заснований на статистичних даних, є науковим припущенням і називається **статистичною гіпотезою**. Статистичні гіпотези підлягають перевірці, мета якої - визначити, чи не суперечить висунута гіпотеза вихідному статистичному матеріалу (вибірці).

Основну гіпотезу, сформульовану в результаті обробки статистичного матеріалу, називають **нульовою гіпотезою і** позначають H_0. На противагу нульовій гіпотезі призначають одну або декілька альтернативних (конкуруючих) гіпотез. Їх позначають H_1, H_2, … і т.д.

Наприклад, якщо перевіряється гіпотеза про рівність параметра a деякому заданому значенню a_0, то як альтернативні гіпотези можна розглянути гіпотези, що a більше або менше a_0:

H_0: $a = a_0$;
H_1: $a > a_0$;
H_2: $a < a_0$;
H_3: $a \neq a_0$.

Вибір альтернативної гіпотези визначається формулюванням задачі.

В якості критеріїв для перевірки статистичних гіпотез використовують випадкові величини (статистики), особливість яких полягає в тому, що кожна з них має свій закон розподілу, що не залежить від закону розподілу генеральної сукупності і вибірки, а залежить від умов обробки вибіркових даних. Значення цих випадкових величин, позначимо їх Z, з відповідними їм імовірностями наводяться в довідкових таблицях.

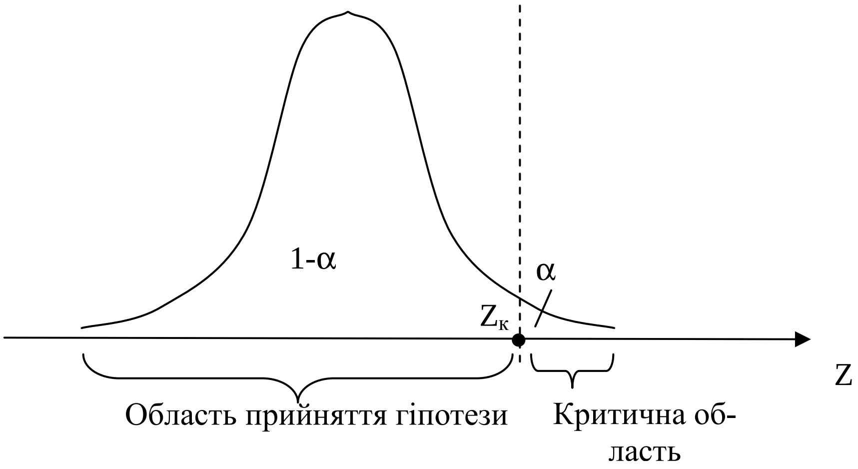


Рис. 10.1 - Застосування критерію для перевірки статистичних гіпотез

На підставі вибіркових даних знаходять значення критерію Z і порівнюють його з табличним значенням, що відповідає умовам обробки даних. Перевірка статистичної гіпотези заснована на принципі, відповідно до якого малоімовірні події вважаються неможливими, а події, що мають велику імовірність, - достовірними. Якщо імовірність розрахункового значення критерію Z досить велика, тобто факт цілком імовірний, то говорять, що гіпотеза не суперечить даним спостереження. Якщо ж ця імовірність мала, тобто подія практично неможлива, то говорять, що нульова гіпотеза суперечить даним спостереження, і її відкидають.

Питання про те, яку імовірність слід вважати досить великою або малою, вирішується не з математичних міркувань, а залежить від наслідків того, що прийнята гіпотеза виявиться невірною. Мала імовірність, при якій значення критерію вважається практично неможливим, позначається α і називається **рівнем значущості**. В практичних задачах звичайно призначають рівень значущості α = 0,05-0,15. Область значень критерію Z, що відповідають рівневі значущості α, називають **критичною областю**. Область значень критерію Z, що відповідають імовірності 1-α, називають **областю прийняття гіпотези**. Значення критерію, яке відокремлює область прийняття гіпотези від критичної області називається **критичною точкою z_k**.

Залежно від того, як сформульовані конкуруючі гіпотези, критична область може бути однобічною (лівосторонньою або правобічною) і двосторонньою. Відповідно критерій може мати одну або дві критичні точки.

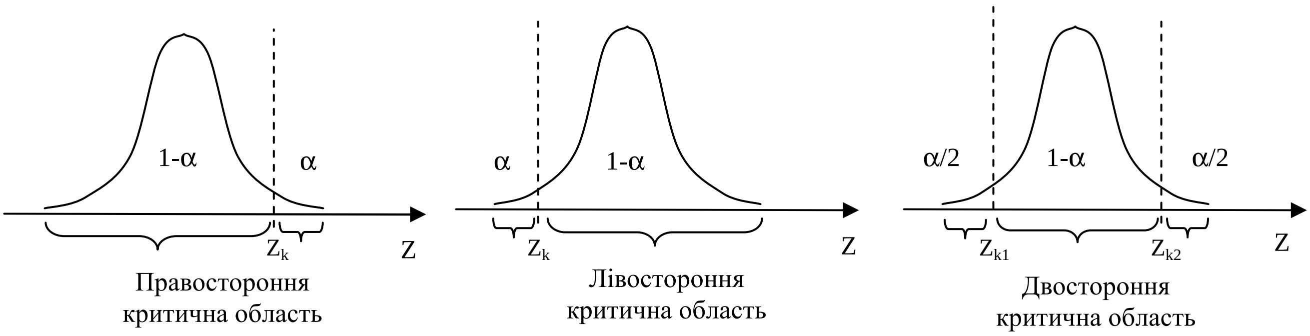


Рис. 10.2 - Розташування критичних точок

Таким чином, перевірка гіпотези заснована на факті, що критерій прийняв значення, імовірність якого може бути більшою або меншою α. Зазначимо, що імовірність даного факту не дорівнює одиниці, а виходить, він не є достовірною подією. Таким чином, результат перевірки гіпотези може виявитися помилковим. Прийнято розрізняти помилки двох видів:

1) помилка 1-го роду - відкинута нульова гіпотеза, в той час як вона була правильною;
2) помилка 2-го роду - прийнята нульова гіпотеза, в той час як вона була невірною.

Якщо перевірка гіпотези показала, що вона не узгоджується з вибірковими даними і повинна бути відкинутою, а задача все-таки вимагає розв'язання, то для цього переглядають вирішення задачі, використовують іншу вибірку з генеральної сукупності або збільшують обсяг вибірки. Тобто проблема все-таки може бути вирішена.

Гірше, якщо зроблено помилку другого роду, тобто, прийнята невірна гіпотеза. Імовірність помилки 2-го роду позначається β. Ця імовірність повинна бути як можна меншою. Тоді імовірність того, що помилку 2-го роду не буде зроблено, визначиться як $1-\beta$. Величина імовірності β залежить від якості використовуваного для перевірки гіпотези критерію. Імовірність $1-\beta$ називається **потужністю критерію**, чим вона більше, тим краще використовуваний критерій, тобто вище надійність перевірки.

Визначимо важливу обставину: статистичні критерії не можуть довести ні однієї гіпотези, вони можуть тільки вказати на відсутність підстав для її спростування.

Розглянемо чотири критерії (нормальний розподіл, t-критерій Стьюдента, F-критерій Фішера і χ^2-критерій Пірсона), які використовуються найбільше часто. Який з названих критеріїв слід використовувати, залежить від характеру розв'язуваної задачі, тобто від формулювання нульової гіпотези H_0. Розглянемо ряд типових задач.

Порівняння вибіркової середньої і генеральної середньої нормальної сукупності

Нехай з нормальної генеральної сукупності взята вибірка обсягом n і визначена вибіркова середня $\tilde{X}$. Припускаємо, що генеральна середня дорівнює $\overline{X}$ і генеральна дисперсія дорівнює $\sigma^2_{ген}$. Необхідно перевірити, чи значуща розбіжність між вибірковою і генеральною середніми, або вона визначена випадковими причинами, тобто незначуща. Запишемо нульову гіпотезу, врахуємо при цьому, що математичне сподівання вибіркової середньої дорівнює генеральній середній:

$$H_0:\ M[\tilde{X}] = \overline{X}.$$

Сформулюємо альтернативну гіпотезу:

$$H_1:\ M[\tilde{X}] \neq \overline{X}.$$

При такому формулюванні альтернативної гіпотези слід побудувати двосторонню критичну область, імовірність влучення в яку дорівнює рівню значущості α. Як мірою розбіжності між вибірковою і генеральною середніми скористаємось випадковою величиною Z:

$$Z = \frac{\tilde{X} - \overline{X}}{\sigma_z}, \tag{10.1}$$

яка є нормованою нормальною випадковою величиною з параметрами $M[Z]=0$ і $\sigma_z = 1$.

Найбільша потужність критерію досягається, якщо імовірність влучення критерію Z у кожен з двох інтервалів критичної області дорівнює $\alpha/2$:

$$P\{|Z|>z_{кр}\}=\alpha/2. \tag{10.2}$$

Оскільки розподіл критерію Z симетричний відносно нуля, критичні точки також розташовані симетрично відносно нуля, тобто область прийняття нульової гіпотези $(-z_{кр}, z_{кр})$.

Для визначення критичних точок можна скористатися функцією Лапласа $\Phi(x)$, що являє собою імовірність влучення нормованої випадкової величини в інтервал $(0,x)$, тобто

$$P\{0<X<x\} = \Phi(x).$$

Оскільки розподіл Z симетричний відносно нуля, за теоремою додавання імовірностей маємо:

$$P\{0< Z <z_{кр}\} + P\{z_{кр} < Z < \infty\}= 1/2,$$

або, виразивши імовірність через функцію Лапласа, дістанемо:

$$\Phi(z_{кр})+ \alpha/2 = ½,$$

звідки

$$\Phi(z_{кр}) = \frac{1-\alpha}{2}. \tag{10.3}$$

Таким чином, знайшовши значення функції Лапласа, можна з таблиці знайти значення її аргументу, тобто критичну точку $z_{кр}$. Тоді двостороння критична область визначається двома нерівностями:

$$Z < -z_{кр};\quad Z > z_{кр},$$

а область прийняття гіпотези:

$$|Z|<z_{кр}. \tag{10.4}$$

Приклад 10.1. Нехай на підставі вибірки обсягом n=15 з нормальної генеральної сукупності отримане значення вибіркової середньої $\tilde{X}$ =100,2. Відомі параметри генеральної сукупності: $\overline{X}$=100 і $\sigma^2_{ген}$ =0,1. Перевірити, чи викликана розбіжність між вибірковими і генеральною середніми випадковими причинами.

Розв'язання: Сформулюємо нульову і альтернативну гіпотези:

$$H_0:\ \tilde{X} = \overline{X};$$

$$H_1:\ \tilde{X} \neq \overline{X}.$$

Задамося п’ятивідсотковим рівнем значущості $\alpha = 0,05$ і знайдемо значення функції Лапласа:

$$\Phi(z_{кр}) = \frac{1-\alpha}{2} = \frac{1-0,05}{2} = 0,475,$$

звідки $z_{кр} = 1,96$.

Знайдемо спостережуване значення критерію:

$$Z = \frac{\tilde{X} - \overline{X}}{\sigma^2_{ген} / \sqrt{n}} = \frac{100,2-100}{(0,1)^2 / \sqrt{15}} = 5,16\,.$$

Таким чином, значення Z=5,16 потрапляє в критичну область, тобто в область відхилення гіпотези H_0, вибіркова середня не погоджується з нульовою гіпотезою. Її варто відхилити, тому що розбіжність між вибірковою і генеральною середніми є невипадковою.

У випадку, коли генеральна дисперсія невідома, в якості критерію використовують t-розподіл (розподіл Стьюдента) з (n-1) ступенями свободи. Спостережуване значення критерію при цьому обчислюють за формулою

$$T = \frac{\tilde{X} - \overline{X}}{\sigma_{выб} / \sqrt{n}}, \tag{10.5}$$

де $\sigma_{выб}$ – вибіркове середнє квадратичне відхилення.

У практичних задачах часто виникає ситуація, коли при визначенні вибіркової середньої відома величина припустимої помилки δ. Завдання полягає у визначенні обсягу вибірки, що забезпечує задану припустиму величину помилки $\delta = \tilde{X} - \overline{X}$. Обсяг вибірки n можна визначити, скориставшись формулою (10.5). Нехай в результаті перевірки нульової гіпотези визначена двостороння критична область, що відповідає рівню значущості α, тоді

$$n = \frac{t^2_{двуст}(\alpha,k) * \sigma^2_{выб}}{(\tilde{X} - \overline{X})^2} = \frac{t^2_{двуст}(\alpha,k) * \sigma^2_{выб}}{\delta^2},$$

де k - число ступенів свободи, рівне n-1.

Приклад 10.2. Нехай на підставі вибірки обсягом n=10 отримане значення вибіркової середньої $\tilde{X}$ =1440 і вибіркове середнє квадратичне відхилення $\sigma_{виб}$=90. Відома генеральна середня $\overline{X}$=1500. Перевірити, чи випадкова розбіжність між вибірковою і генеральною середніми.

Розв'язання: Запишемо нульову гіпотезу:

$$H_0: M[\tilde{X}] = a.$$

Сформулюємо альтернативну гіпотезу:

$$H_1: M[\tilde{X}] \neq \overline{X}.$$

Оскільки генеральна дисперсія не відома, для перевірки скористаємося t-розподілом з 10-1=9 ступенями свободи, що відповідає п'ятивідсотковому рівню значущості α=0,05. З таблиці значень t-критерію Стьюдента знаходимо $t_{0,05/9}=\pm 2{,}26$. Розрахуємо спостережуване значення критерію:

$$T = \frac{\tilde{X} - \overline{X}}{\sigma_{выб} / \sqrt{n}} = \frac{1440 - 1500}{90 / \sqrt{10}} = -2{,}11.$$

Таким чином, спостережуване значення критерію T=-2,11 лежить правіше табличного t=-2,26, звідки видно, що імовірність вибіркової середньої $\tilde{X}$ =1440, отриманої на підставі 10 одиниць, взятих з нормальної генеральної сукупності з генеральною середньою $\overline{X}$=1500 становить більше 5%. Таким чином, нульова гіпотеза погоджується з вибіркою, її можна прийняти.

Порівняння двох дисперсій нормальних генеральних сукупностей

Нехай є дві незалежні вибірки і треба визначити, чи взяті вони з нормальних генеральних сукупностей X і Y з однаковою дисперсією. Запишемо нульову гіпотезу:

$$H_0: D[X] = D[Y], \tag{10.6}$$

і сформулюємо альтернативну гіпотезу:

$$H_1: D[X] \neq D[Y]. \tag{10.7}$$

Як критерій для перевірки нульової гіпотези про рівність дисперсій нормальних генеральних сукупностей приймають випадкову величину, що являє собою відношення більшої дисперсії до меншої:

$$F = \frac{\sigma_1^2}{\sigma_2^2}, \tag{10.8}$$

де $\sigma_1 > \sigma_2$.

Нульова гіпотеза передбачає, що дві вибірки незалежні й взяті з нормальних генеральних сукупностей з однаковими дисперсіями, в цьому випадку F=1. Однак, навіть якщо гіпотеза вірна, то малоімовірно, що σ_1 прикмет точно таке ж значення, як σ_2 через вплив випадковості. Таким чином, завдання полягає в тому, щоб перевірити, чи буде випадкова величина F досить близька до одиниці.

Випадкова величина F має розподіл Фішера, він залежить тільки від ступенів свободи $k_1=n_1-1$ і $k_2=n_2-1$, де n_1 і n_2 обсяг вибірки з більшою і меншою вибірковими дисперсіями відповідно, і не залежить від інших параметрів.

Для перевірки нульової гіпотези (10.6) при конкуруючій гіпотезі (10.7) будують двосторонню критичну область, що відповідає рівню значущості α. Двостороння критична область визначається двома нерівностями:

$$F < F_{кр1}; \qquad F > F_{кр2},$$

а область прийняття гіпотези:

$$F_{кр1}<F<F_{кр2},$$

причому імовірність влучення критерію в кожний з двох інтервалів критичної області дорівнює $\alpha/2$.

Приклад 10.3. З двох вибірок з обсягом $n_1=10$ і $n_2=8$ відповідно визначені вибіркові дисперсії, значення яких $\sigma^2_{виб1} = 11{,}45$ і $\sigma^2_{виб2} = 58{,}26$. Перевірити, чи взяті дві вибірки з нормальних генеральних сукупностей з однаковими дисперсіями σ^2_1 і σ^2_2.

Розв'язання. Запишемо нульову і альтернативну гіпотези:

$$H_0: \sigma^2_1 = \sigma^2_2,$$

$$H_1: \sigma^2_1 \neq \sigma^2_2.$$

Будемо перевіряти нульову гіпотезу за допомогою F-критерію з двома межами, на 10%-му рівні значущості, що рівнозначне 5%-му рівню значущості з однією межею. Для встановлення критичного значення F задамося $\alpha=0{,}05$.

Розрахуємо значення F-критерію

$$F = \frac{58{,}26}{11{,}45} = 5{,}09$$

і знайдемо критичне значення F-критерію, з огляду на те, що число ступенів свободи для більшої дисперсії дорівнює 7, а для меншої - 9, тоді дістанемо F= 3,29.

Область прийняття гіпотези визначається співвідношенням

$$F_{кр1}<F<F_{кр2}.$$

Праву критичну точку знаходимо безпосередньо з таблиці розподілу Фішера $F_{кр2}(\alpha/2;k_1;k_2)=3{,}29$. У випадку конкуруючої гіпотези $\sigma^2_1\neq\sigma^2_2$ для визначення двосторонньої критичної області досить визначення однієї правої критичної точки. Так, імовірність влучення в неї дорівнює 0,05, а імовірність влучення у двосторонню критичну область 0,05*2=0,1.

Таким чином, $F=5{,}09 >F_{кр}=3{,}29$. Це є підставою відкинути нульову гіпотезу, отже дві вибірки взяті з нормальних генеральних сукупностей з різними дисперсіями σ^2_1 і σ^2_2.

Порівняння вибіркової і генеральної дисперсій нормальної сукупності

Нехай з генеральної сукупності взята вибірка обсягу n і визначена вибіркова дисперсія $\sigma^2_{виб}$. Потрібно перевірити нульову гіпотезу, яка полягає в тому, що генеральна дисперсія дорівнює гіпотетичному значенню σ^2_0. Нульова гіпотеза в цьому разі записується в такий спосіб:

$$H_0: M[\sigma^2_{виб}] = \sigma^2_0.$$

Інакше кажучи, потрібно встановити, значущо або незначущо розрізняються виправлені вибіркова і гіпотетична генеральна дисперсії.

Як критерій для перевірки нульової гіпотези використовується випадкова величина

$$\frac{(n-1)\sigma^2_{виб}}{\sigma^2_0}.$$

Доведено, що ця випадкова величина має розподіл χ^2 з n-1 ступенями свободи:

$$\chi^2 = \frac{(n-1)\sigma^2_{виб}}{\sigma^2_0}. \qquad (10.9)$$

У випадку, коли альтернативна гіпотеза H_0: M[$\sigma^2_{виб}$] $\neq \sigma^2_0$, будується двостороння критична область, імовірність влучення в яку дорівнює прийнятому рівню значущості α. Межі критичної області знаходять виходячи з того, щоб імовірність влучення критерію χ^2 у лівий і правий інтервали критичної області дорівнювала $\alpha/2$, тобто

$$\begin{aligned} P\{\chi^2 < \chi^2_{лів}\} &= \frac{\alpha}{2} \\ P\{\chi^2 > \chi^2_{прав}\} &= \frac{\alpha}{2} \end{aligned}.$$

Праву межу критичної області можна визначити безпосередньо з таблиці, а для визначення лівої користуються тим, що сума імовірностей протилежних подій дорівнює одиниці:

$$P\{\chi^2 < \chi^2_{лів}\} + P\{\chi^2 > \chi^2_{лів}\} = 1,$$

звідки

$$P\{\chi^2 > \chi^2_{лів}\} = 1 - P\{\chi^2 < \chi^2_{лів}\} = 1 - \frac{\alpha}{2}.$$

Таким чином, ліву критичну точку визначають з таблиці, виходячи з того, щоб імовірність влучення критерію χ^2 правіше цієї точки дорівнювала $1-\alpha/2$.

Якщо спостережуване значення критерію χ^2 лежить у межах

$$\chi^2_{лів} \leq \chi^2_{спост} \leq \chi^2_{прав},$$

відкидати нульову гіпотезу немає підстав. Якщо $\chi^2_{спост} \leq \chi^2_{лів}$ або $\chi^2_{спост} \geq \chi^2_{прав}$, нульову гіпотезу відкидають.

Приклад 10.4. На підставі вибірки обсягом n=13 з нормальної генеральної сукупності визначена вибіркова дисперсія $\sigma^2_{виб}$ = 10,3. Гіпотетичне значення генеральної дисперсії σ^2_0 = 12. Перевірити нульову гіпотезу H_0: σ^2_0 = 12 при альтернативній гіпотезі H_1: $\sigma^2_0 \neq 12$.

Розв'язання. Знайдемо розрахункове значення критерію χ^2 за формулою (10.9)

$$\chi^2 = \frac{(n-1)\sigma^2_{виб}}{\sigma^2_0} = \frac{(13-1)*10{,}3}{12} = 10{,}3.$$

Задамося рівнем значущості α=0,1 і визначимо двосторонню критичну область з урахуванням того, що число ступенів свободи 13-1=12. З таблиці знаходимо $\chi^2_{лів} = 5{,}23 \le \chi^2_{спост} = 10{,}3 \le \chi^2_{прав} = 21{,}03$, таким чином, немає підстав відкинути нульову гіпотезу. Інакше кажучи, вибіркова дисперсія $\sigma^2_{виб}$ =10,3 і гіпотетична генеральна дисперсія $\sigma^2_0 = 12$ розрізняються незначуще.

Перевірка гіпотези про нормальний розподіл генеральної сукупності. Критерій згоди χ^2

Якщо за вибіркою спостережень визначався закон розподілу генеральної сукупності, то виникає необхідність оцінити розбіжність між емпіричним і теоретичним розподілами. Для цього використовують ***критерії згоди***, які дозволяють судити, якою є розбіжність між **емпіричним** і **теоретичним** розподілами - випадковою або значущою.

Якщо розбіжність виявиться випадковою, то вважають, що дані спостережень (вибірки) не суперечать висунутій гіпотезі про закон розподілу генеральної сукупності і, отже, гіпотезу приймають; якщо ж розбіжність виявиться значущою, то дані спостережень суперечать гіпотезі, і її відкидають. В якості міри розбіжності використовують деяку величину U. Є кілька критеріїв згоди: критерій χ^2 (Пірсона), критерій Колмогорова, критерій Романовського та ін.

Пірсон запропонував використовувати як міру розбіжності суму квадратів відхилень ($p_i^* - p_{ip}$), взятих з деякими вагами c_i, тоді

$$U = \sum_{i=1}^{k} c_i (p_i^* - p_{ip})^2 , \qquad (10.10)$$

де p_i^* – частота появи ознаки X на i-му інтервалі; p_{ip} - теоретична імовірність тієї ж події; k - число інтервалів; c_i – ваговий коефіцієнт, який враховує, що відхилення ($p_i^* - p_{ip}$), які належать до різних груп ряду, не можна вважати рівноправними за значущостю, оскільки те саме за абсолютною величиною відхилення може бути малим для великого p_i і істотним для малого p_i.

Пірсон показав, що коли прийняти

$$c_i = \frac{n}{p_i}, \qquad (10.11)$$

то величина U при збільшенні числа дослідів n наближається до величини χ^2, розподіл якої залежить тільки від числа ступенів свободи:

$$r = k - s, \qquad (10.12)$$

де k - число інтервалів; s - число зв'язків (число незалежних умов).

Підставивши (10.11) в (10.10) і врахувавши, що $p_i^* = \frac{m_i}{n}$, дістанемо:

$$\chi^2_P = \sum_{i=1}^{k} \frac{(m_i - np_{ip})^2}{np_{ip}} \qquad (10.13)$$

Чим більше узгоджуються емпіричний і теоретичний розподіли, тим менше розрізняються емпіричні й теоретичні частоти і тем менше значення χ^2_P.

Звідси випливає, що χ^2_P характеризує близькість емпіричного і теоретичного розподілів.

Існують довідкові таблиці, в яких указана імовірність того, що в результаті впливу випадкових факторів величина χ^2 прийме значення не менше обчисленого за даними вибірки.

Для визначеності приймають рівень значущості α (звичайно α = 0,1-0,15). Якщо імовірність, знайдена з таблиць, виявиться менше α, то це означає, що в результаті впливу випадкових причин наступила подія, яка практично неможлива. Таким чином, той факт, що χ^2 прийняла значення χ^2_P не можна пояснити випадковими причинами; його можна пояснити тим, що генеральна сукупність не розподілена за передбачуваним законом розподілу й, виходить, висунута гіпотеза про закон розподілу генеральної сукупності повинна бути відкинута. Якщо імовірність, знайдена з таблиць, перевищує α, то гіпотеза про закон розподілу генеральної сукупності погоджується з даними спостережень і тому може бути прийнята.

Приклад 10.5. У прикладі 8.2 визначені параметри нормального розподілу помилки приладу на підставі 500 вимірювань. Результати вимірювань зведені в групований варіаційний ряд:

групи	-4 ...-3	-3 ...-2	-2 ...-1	-1 ...0	0 ...1	1 ...2	2 ...3	3 ...4
m_i	6	25	72	133	120	88	46	10

Перевірити гіпотезу про нормальний розподіл помилки X з параметрами $\tilde{X} = 0{,}168$ і $\sigma_{виб} = 1{,}448$.

Розв'язання. Для перевірки гіпотези про нормальний розподіл помилки X скористаємося критерієм згоди Пірсона χ^2. Розрахункове значення критерію визначаємо за формулою (10.13), де p_{ip} – імовірності влучення X в кожен з 8 інтервалів, обчислені з аналітичного виразу

$$f^*(x) = \frac{1}{1{,}448\sqrt{2\pi}} \exp\left\{-\frac{(x-0{,}168)^2}{2*2{,}126}\right\}.$$

Попередньо знайдемо їхні значення, скориставшись формулою

$$p_{ip} = P\{x_i < X < x_{i+1}\} = F(x_{i+1}) - F(x_i),$$

де F(x) – нормальна функція розподілу; x_i – межі інтервалів.

Значення нормальної функції розподілу знайдемо за допомогою статистичної функції Microsoft Excel НОРМРАСП. Функція НОРМРАСП повертає нормальну функцію розподілу для зазначеного середнього і стандартного відхилення (рис. 10.3). Синтаксис функції:

НОРМРАСП(х;среднее;стандартное_откл;интегральная),

де х — значення, для якого будується розподіл; среднее — середнє арифметичне нормального розподілу; стандартное_откл — стандартне відхилення нормального розподілу; интегральная — логічне значення, що визначає форму функції. Якщо интегральная має значення ИСТИНА, то функція НОРМРАСП повертає інтегральну функцію розподілу; якщо цей аргумент має значення ЛОЖЬ, то повертається функція щільності розподілу.

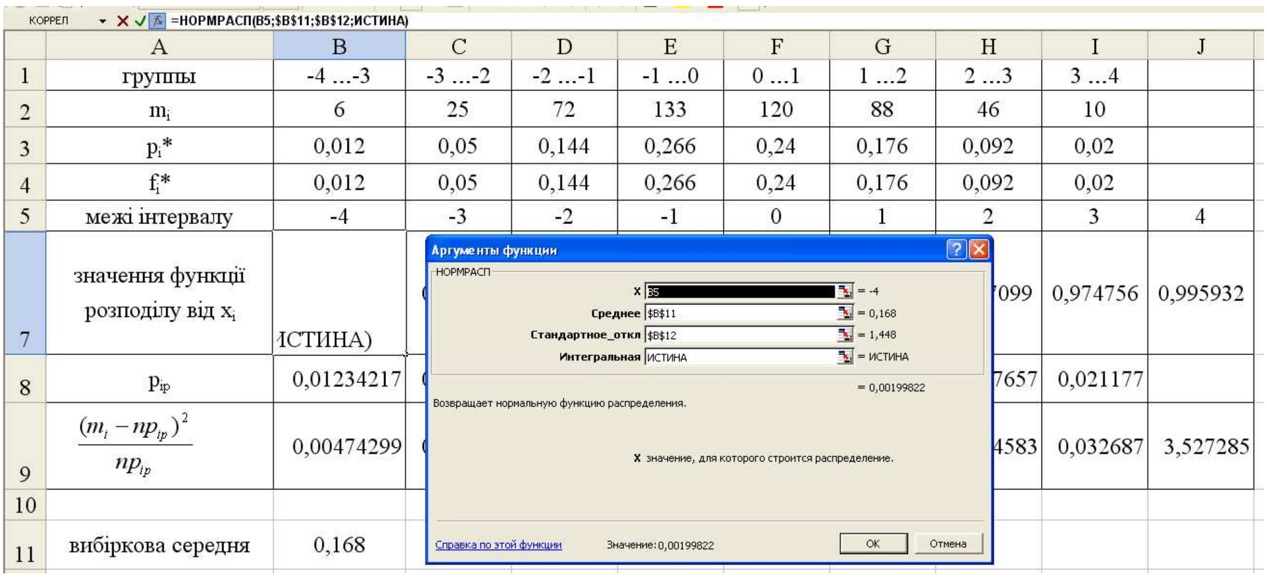


Рис. 10.3 - Введення аргументів функції НОРМРАСП

Результати розрахунків зведемо в таблицю

групи	-4 ...-3	-3 ...-2	-2 ...-1	-1 ...0	0 ...1	1 ...2	2 ...3	3 ...4	
m_i	6	25	72	133	120	88	46	10	
p_i*	0,012	0,05	0,144	0,266	0,24	0,176	0,092	0,02	
f_i*	0,012	0,05	0,144	0,266	0,24	0,176	0,092	0,02	
межі інтервалу	-4	-3	-2	-1	0	1	2	3	4
значення функції розподілу від x_i	0,002	0,0143	0,067	0,21	0,454	0,717	0,897	0,975	0,996
p_{ip}	0,0123	0,0528	0,143	0,244	0,263	0,18	0,078	0,021	
$\frac{(m_i - np_{ip})^2}{np_{ip}}$	0,0047	0,0756	0,005	1,003	1,04	0,042	1,325	0,033	3,53

Таким чином, знайдене розрахункове значення χ^2=3,527.

Табличні значення випадкової величини наводяться в довідковій таблиці, входом в яку є рівень значущості α і число ступенів свободи. Задамося рівнем значущості α = 0,05. Число ступенів свободи визначимо як різницю між числом інтервалів, на які потрапляють експериментальні значення випадкової величини X, їх вісім, і числом зв'язків, що визначається умовами задачі. У процесі розв'язання ми ставили умови:

$$\sum_{i=1}^{8} p_i^* = 1; \qquad \tilde{X} = \sum_{i=1}^{8} x_i p_i^* = 0{,}168; \qquad \sigma_{виб}^2 = \tilde{\alpha}_2 - \tilde{X}^2 = 1{,}448,$$

отже число зв'язків дорівнює трьом. Тоді число ступенів свободи

$$8-3=5.$$

Табличні значення χ^2 зручно знаходити з використанням статистичних функцій Microsoft Excel ХИ2РАСП і ХИ2ОБР. Функція ХИ2РАСП повертає імовірність розподілу χ^2. Синтаксис функції:

ХИ2РАСП(x;степени_свободы),

де x — це значення, для якого потрібно обчислити розподіл.
Функція ХИ2РАСП обчислює імовірність P(X> x), де x — випадкова величина χ^2.

Функція ХИ2ОБР повертає значення, зворотне до імовірності розподілу χ^2. Якщо імовірність = ХИ2РАСП(x;...), то ХИ2ОБР(імовірність;...) = x. Синтаксис функції:

ХИ2ОБР(вероятность;степени_свободы),

де вероятность — це імовірність, пов'язана з розподілом χ^2.

У нашому прикладі байдуже, яку із зазначених функцій використовувати. Знаючи розрахункове значення χ^2 і число ступенів свободи, можна скористатися функцією ХИ2РАСП (рис. 10.4) і дістати імовірність того, що розрахункове значення $\chi^2 = 3{,}527$, не перевищить табличне значення $\chi^2_{табл.}$ Оскільки отримана імовірність більше заданого рівня значущості

$$p=0{,}6 > \alpha=0{,}05,$$

підстав щоб відкинути гіпотезу про нормальний розподіл випадкової величини X немає, а неузгодженість між статистичним і теоретичним розподілами викликана випадковими причинами.

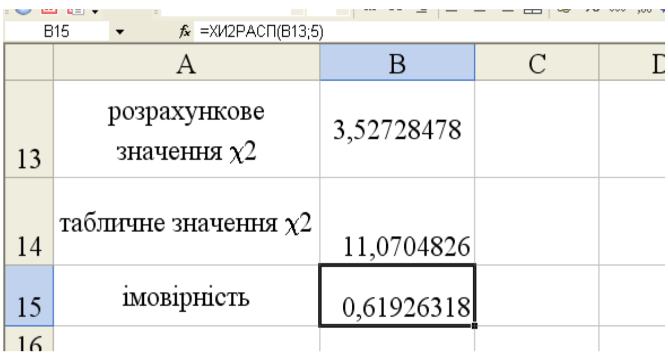


Рис. 10.4 - Використання функції ХИ2РАСП

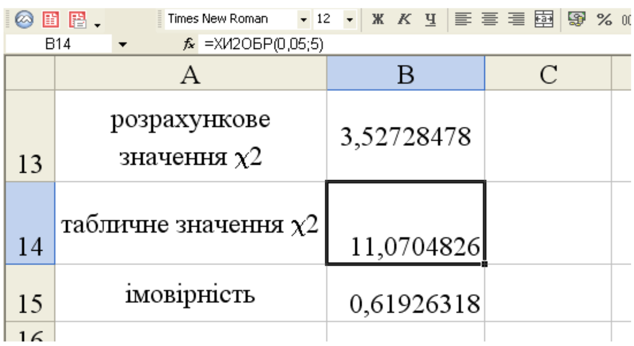


Рис. 10.5 - Використання функції ХИ2ОБР

Знаючи рівень значущості і число ступенів свободи, можна скористатися функцією ХИ2ОБР (рис. 10.5) і дістати табличне значення $\chi^2_{табл}=11{,}07$. Оскільки розрахункове значення χ^2 менше табличного

$$\chi^2 = 3{,}527<\chi^2_{табл}= 11{,}07$$

можна вважати, що неузгодженість між статистичним і теоретичним розподілами викликана випадковими причинами і немає підстав відкинути гіпотезу про нормальний розподіл помилки X з параметрами $\tilde{X} = 0{,}168$ і $\sigma_{выб} = 1{,}448$.

Приклад 10.6. Зроблено 800 спостережень над випадковою величиною X. Результати спостережень подани у вигляді таблиці:

x_i	0	1	2	3	4	5	6	7	8	9	10
m_i	25	81	124	146	175	106	80	35	16	6	6

Потрібно оцінити правдоподібність гіпотези, яка полягає в тому, що X розподілена за законом Пуассона. Як рівень значущості прийняти α=0,15.

Розв'язання. Для перевірки гіпотези скористаємося критерієм згоди Пірсона χ^2. Розрахункове значення критерію визначається за формулою (10.13), де p_{ip} – імовірність, знайдена за теоретичною залежністю, тобто за виразом закону Пуассона:

$$P(k) = \frac{(\lambda\tau)^k}{k!} e^{-\lambda\tau},$$

де $\lambda\tau$ - математичне сподівання випадкової величини X.

Для розрахунку імовірностей p_{ip} за цією формулою скористаємося замість математичного сподівання випадкової величини X вибірковою середньою $\tilde{X}$, котру визначимо на підставі статистичних даних. Для цього знайдемо статистичні частоти p_i*= m_i/800. Результати розрахунку розмістимо в таблиці

x_i	0	1	2	3	4	5	6	7	8	9	10
m_i	25	81	124	146	175	106	80	35	16	6	6
p_i*	0,031	0,101	0,155	0,183	0,219	0,133	0,1	0,044	0,02	0,008	0,008
$x_i p_i$*	0	0,101	0,31	0,548	0,875	0,663	0,6	0,306	0,16	0,068	0,075

Вибіркову середню визначимо за формулою $\tilde{X} = \sum_{i=1}^{11} x_i p_i$, одержимо $\tilde{X}$ =3,705. Тепер можна визначити p_{ip}. Скористаємося для цього замість формули Пуассона статистичною функцією ПУАССОН (рис. 10.6). Функція повертає розподіл Пуассона. Синтаксис функції:

ПУАССОН(x;среднее;интегральная),

де x — число подій (k); средняя — математичне сподівання; интегральная — логічне значення, що визначає форму розподілу імовірностей, що повертається. Якщо аргумент «интегральная» має значення ИСТИНА, то функція ПУАССОН повертає інтегральний розподіл Пуассона, тобто імовірність того, що число випадкових подій буде від 0 до x включно. Якщо цей аргумент має значення ЛОЖЬ, то повертається імовірність того, що подій буде в точності x.

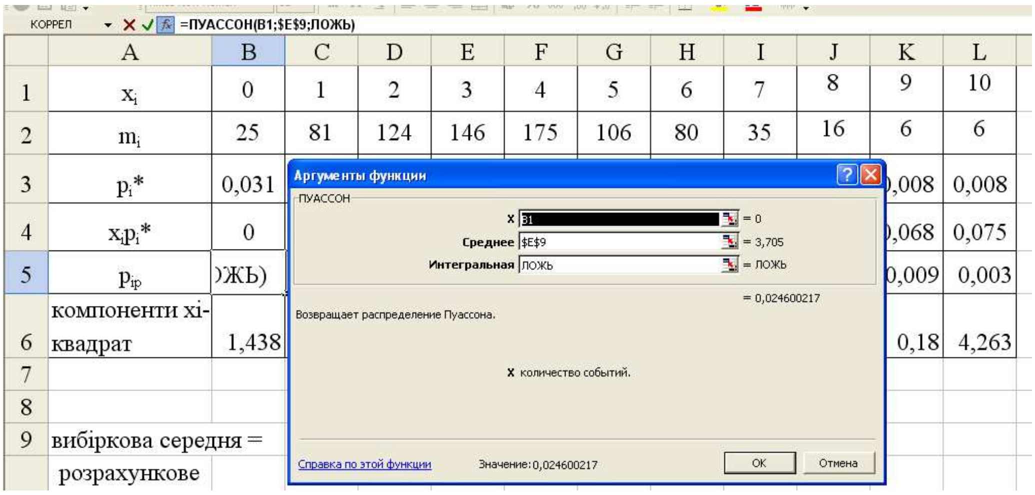


Рис. 10.6 - Введення аргументів статистичної функції ПУАССОН

Визначимо компоненти χ^2 і занесемо їх у таблицю

x_i	0	1	2	3	4	5	6	7	8	9	10
m_i	25	81	124	146	175	106	80	35	16	6	6
p_i*	0,031	0,101	0,155	0,183	0,219	0,133	0,1	0,044	0,02	0,008	0,008
$x_i p_i$*	0	0,101	0,31	0,548	0,875	0,663	0,6	0,306	0,16	0,068	0,075
p_{ip}	0,025	0,091	0,169	0,209	0,193	0,143	0,088	0,047	0,022	0,009	0,003
компо-ненти хі-квадрат	1,44	0,89	0,91	2,6	2,72	0,63	1,22	0,16	0,10	0,18	4,26

Отримали χ^2 = 15,11. Число зв'язків у цьому випадку визначимо з умов, що накладаються: Σpі = 1 $\tilde{X} = \sum_{i=1}^{11} x_i p_i$, тобто з умов збігу гіпотетичних параметрів зі статистичними. Таким чином, число зв'язків дорівнює двом.

Застосовуючи розрахункове значення χ^2 = 15,11 і число ступенів свободи 11-2=9 за допомогою статистичної функції ХИ2РАСП дістанемо імовірність р=0,088. Порівнявши її з рівнем значущості р=0,088<α=0,15, зробимо висновок, що гіпотеза про розподіл Х за законом Пуассона суперечить статистичним даним і повинна бути відкинута.

Приклад 10.7. При вісьмох підкиданнях монети герб з'явився три рази. Чи можна вважати монету симетричною? (Перевірити гіпотезу про те, що число появ герба має біноміальний розподіл з параметром р=0,5).

Розв'язання. Число появ герба в досліді підкидання однієї монети розподілено за біноміальним законом з імовірністю появи герба в одному досліді, рівною 0,5.

$$P_n(k) = C_n^k * p^k (1-p)^{n-k},$$

де C_n^k - число сполучень із n елементів по k; p – імовірність появи герба при кожному підкиданні монети.

Для перевірки гіпотези скористуємося критерієм згоди Пірсона χ^2. Розрахункове значення критерію знаходимо за формулою (10.13)

$$\chi_P^2 = \sum_{i=1}^{k} \frac{(m_i - np_{ip})^2}{np_{ip}},$$

де p_{ip} – імовірність, знайдена з теоретичної залежності, тобто з виразу біноміального закону розподілу.

Визначимо імовірність, що при вісьмох підкиданнях монети герб з'явиться не більше трьох разів

$$C_8^0 * 0{,}5^0 0{,}5^{8-0} + C_8^1 * 0{,}5^1 0{,}5^{8-1} + C_8^2 * 0{,}5^2 0{,}5^{8-2} + C_8^3 * 0{,}5^3 0{,}5^{8-3} =$$

$$= 1 * 0{,}0039 + 8 * 0{,}0039 + 28 * 0{,}0039 + 56 * 0{,}0039 = 0{,}3632.$$

Тепер знайдемо значення χ^2

$$\chi_P^2 = \frac{(3 - 8 * 0{,}3632)^2}{8 * 0{,}3632} = 0{,}00307.$$

Число ступенів свободи дорівнює 1-0=1, тому що число інтервалів дорівнює 1, а число зв'язків дорівнює 0. Скористаємося функцією ХИ2РАСП і дістанемо імовірність p=0,95. Оскільки імовірність отриманого значення χ^2 = 0,00307 велика, передбачувана гіпотеза про біноміальний розподіл числа появ герба при підкиданні монети не може бути відкинута. Підкидали, швидше за все, симетричну монету.

Елементи дисперсійного аналізу

Дисперсійний аналіз - метод математичної статистики, застосовуваний для аналізу результатів спостережень, які залежать від різних одночасно діючих факторів. Завдання дисперсійного аналізу - вибір найбільш важливих факторів, оцінка їхнього впливу і т.п.

Метод дисперсійного аналізу розроблений англійським статистиком Р.Фішером. В основі методу лежить порівняння дисперсій. На практиці дисперсійний аналіз застосовують у задачах, де потрібно оцінити вплив деякого фактору F на кількісну ознаку X. Суть дисперсійного аналізу зводиться до порівняння дисперсії, обумовленої впливом фактора F (факторної дисперсії) з дисперсією, обумовленою випадковими причинами (залишковою дисперсією). Очевидно, що коли вплив фактора F є значущим, то і відмінність факторної дисперсії від залишкової дисперсії повинна бути значущою. І навпаки, якщо вплив фактора незначущий, то факторна і залишкова дисперсія відрізняються незначуще.

Нехай значення ознаки X отримані в результаті спостереження p різних груп дослідів із числом спостережень в j-й групі, рівним q. Середнє значення ознаки X в кожній j-й групі (групова середня) знайдемо за формулою:

$$\bar{x}_{грj} = \frac{\sum_{i=1}^{q} x_{ij}}{q}.$$

Результати спостережень зведені в таблицю:

Номер досліду	Номер групи					
	1	2	…	j	…	p
1	x_{11}	x_{12}	…	x_{1j}	…	x_{1p}
2	x_{21}	x_{22}	…	x_{2j}	…	x_{2p}
…	…	…	…	…	…	…
i	x_{i1}	x_{i2}	…	x_{ij}	…	x_{ip}
…	…	…	…		…	
q	x_{q1}	x_{q2}	…	x_{qj}	…	x_{qp}
групова середня	$\bar{x}_{гр1}$	$\bar{x}_{гр2}$	…	$\bar{x}_{грj}$	…	$\bar{x}_{грp}$

Загальне число спостережуваних значень ознаки X дорівнює pq. Загальна середня знаходимо за формулою:

$$\bar{x} = \frac{\sum_{j=1}^{p} \sum_{i=1}^{q} x_{ij}}{pq}.$$

Можна оцінити повне розсіювання ознаки X, викликане як випадковими причинами, так і впливом фактора F, визначивши суму квадратів відхилень всіх спостережуваних значень x_{ij} від загальної середньої. Вона називається **загальною** сумою квадратів відхилень і визначається за формулою

$$S_{общ} = \sum_{j=1}^{p} \sum_{i=1}^{q} (x_{ij} - \bar{x})^2 . \qquad (10.14)$$

Вважають, що фактор F впливає на різні групи значень ознаки. Розсіювання за фактором або розсіювання між групами можна оцінити, визначивши суму квадратів відхилень групових середніх від загальної середньої. Її називають **факторною** сумою квадратів відхилень і визначають за формулою:

$$S_{факт} = q * \sum_{j=1}^{p} (\bar{x}_{грj} - \bar{x})^2 , \qquad (10.15)$$

де $\bar{x}_{грj}$ – середня j-ї групи.

Вважають, що на значення ознаки в j-й групі фактор F впливає однаково, а їхнє розсіювання обумовлене впливом випадкових причин. Суму квадратів відхилень спостережуваних значень ознаки X від своєї груповий середньої $\bar{x}_{грj}$ називають **залишковою** сумою квадратів відхилень. Залишкова сума квадратів відхилень характеризує розсіювання в середині групи і визначається за формулою

$$S_{ост} = \sum_{j=1}^{p} \sum_{i=1}^{q} (x_{ij} - \bar{x}_{грj})^2 \qquad (10.16)$$

Можна показати, що справедливо співвідношення

$$S_{общ} = S_{факт} + S_{ост}, \qquad (10.17)$$

яке найчастіше використовують для визначення залишкової суми квадратів відхилень

$$S_{ост} = S_{общ} - S_{факт} . \qquad (10.18)$$

Оскільки дисперсійний аналіз передбачає порівняння дисперсій, то використовуючи загальну, факторну і залишкову суми квадратів відхилень, визначають відповідні дисперсії.

Загальна дисперсія:

$$s^2_{общ} = \frac{S_{общ}}{pq-1}, \qquad (10.19)$$

де pq-1 = n-1 - число ступенів свободи загальної дисперсії.

Факторна дисперсія:

$$s^2_{факт} = \frac{S_{факт}}{p-1} \qquad (10.20)$$

де p-1 - число ступенів свободи факторної дисперсії; p - число груп впливу фактору F.

Залишкова дисперсія:

$$s^2_{ост} = \frac{S_{ост}}{p(q-1)} \qquad (10.21)$$

де p(q-1) - число ступенів свободи залишкової дисперсії, обумовлене як різниця між числами ступенів свободи загальної і факторної дисперсій:

$$(pq - 1) - (p -1) = p(q-1).$$

Спостережуване значення F-критерію знаходимо за формулою

$$F_{спост} = \frac{S_{факт}}{S_{ост}}. \qquad (10.22)$$

Припустимо, що вплив фактора F відсутній. У цьому разі групові середні $\bar{x}_{грj}$ приймають різні значення в результаті впливу тільки випадкових причин, а виходить, розрізняються незначуще. Відповідно факторна і залишкова дисперсії є незміщеними оцінками невідомої генеральної дисперсії і також розрізняються незначуще. У такій задачі формулюють нульову гіпотезу про рівність факторної і залишкової дисперсій. Якщо порівняти оцінки цих дисперсій за критерієм F, то критерій вкаже, що гіпотезу можна прийняти.

Якщо нульова гіпотеза про рівність групових середніх (а отже факторної і залишкової дисперсій) помилкова, то зі зростанням розбіжності між груповими середніми буде збільшуватися факторна дисперсія і спостережуване значення критерію F. При $F_{спост} > F_{кр}$ нульова гіпотеза про рівність факторної і залишкової дисперсій буде відкинута.

Таким чином, щоб перевірити нульову гіпотезу про рівність групових середніх нормальних сукупностей з однаковими дисперсіями, необхідно перевірити за критерієм F нульову гіпотезу про рівність факторної і залишкової дисперсій. Причому, якщо факторна дисперсія виявиться менше залишкової, то із цього випливає справедливість гіпотези про рівність групових середніх, і F-критерій можна не обчислювати.

Приклад 10.8. Розглянемо виробничу ситуацію. Три майстри проводили перевірку однотипних пристроїв. Кожний з них перевірив різне число пристроїв і виявив різне число дефектів при кожній перевірці. Результати роботи показані в таблиці.

Перевірка (i)	Майстер (j)			Усього
	1	2	3	
1	11	6	8	25
2	7	1	7	15
3	8	2	9	19
4	4	-	4	8
5	5	-	-	5
виявлено дефектів	35	9	28	72
виконано перевірок	5	3	4	12
середнє число дефектів	7	3	7	6

Аналізуючи їхню роботу, керівник висунув припущення про різну кваліфікацію майстрів, тому що на перший погляд середнє число знайдених дефектів при одній перевірці в другого майстра помітно відрізняється. Перевірити гіпотезу про вплив кваліфікації майстрів на результати роботи.

Розв'язання: Виконаємо розрахунки розсіювання результатів. Визначимо факторну суму квадратів відхилень і факторну дисперсію:

$$S_{факт} = q*\sum_{j=1}^{p}(\bar{x}_{грj} - \bar{x})^2 = 5*(7-6)^2+3*(3-6)^2+4*(7-6)^2 = 36;$$

$$s^2_{факт} = \frac{S_{факт}}{p-1} = \frac{36}{3-1} = 18$$

Визначимо залишкову суму квадратів відхилень:

$$S_{ост} = \sum_{j=1}^{p}\sum_{i=1}^{q}(x_{ij} - \bar{x}_{грj})^2 = (11-7)^2+(7-7)^2+(8-7)^2+(4-7)^2+(5-7)^2+(6-3)^2+(1-3)^2+$$

$$+(2-3)^2+(8-7)^2+(7-7)^2+(9-7)^2+(4-7)^2=58.$$

Визначимо число ступенів свободи залишкової дисперсії як різницю числа ступенів свободи загальної і факторної дисперсій:

$$12-3=9$$

і залишкову дисперсію

$$s^2_{ост} = \frac{S_{ост}}{p(q-1)} = \frac{58}{9} = 6{,}44\,.$$

Визначимо спостережуване значення F-критерію:

$$F_{спост} = \frac{18}{6,44} = 2,79.$$

Знайдемо табличне значення F-критерію, що відповідає рівню значущості α=0,05:

$$F_{0,95}(2;9) = 4,26.$$

Таким чином, $F_{0,95}(2;9) = 4,26 > F_{спост} = 2,79$, отже немає підстав відкидати нульову гіпотезу про те, що відмінність між факторною і залишковою дисперсіями, а також між груповими середніми не є значущою. Обвинувачувати другого майстра в низькій кваліфікації немає підстав.

Елементи регресійного аналізу

Отримане за методом найменших квадратів рівняння регресії $\overline{y}=\varphi(x)$ є статистичною гіпотезою, що вимагає перевірки. Параметри цієї залежності, отримані на основі випадкових значень результативної ознаки Y_i, є випадковими величинами. МНК дозволяє визначити лише їхні точкові оцінки, інтервал розподілу яких накриває дійсні значення параметрів з деякою імовірністю. Одним завданням регресійного аналізу є визначення помилки від використання залежності $\overline{y}=\varphi(x)$ для прогнозування і прийняття рішень. Іншим завданням регресійного аналізу є перевірка адекватності рівняння регресії, тобто його відповідності статистичним даним. Це завдання обумовлене, зокрема, тим, що клас апроксимуючої кривої вибирається евристично. До задач регресійного аналізу відносять також оцінку значущості як рівняння регресії в цілому, так і окремих його параметрів.

Оскільки не всі точки поля кореляції лежать на лінії регресії, завжди має місце їхній розкид, обумовлений як впливом фактора X, тобто регресією Y на X, так і інших факторів, не врахованих рівнянням регресії. Перевірку адекватності рівняння регресії експериментальним даним виконують за F-критерієм Фішера. При цьому висувають нульову гіпотезу, що коефіцієнт регресії дорівнює нулю, $\rho_{yx}=0$, тобто що фактор X не має впливу на результативну ознаку Y. Для розрахунку F-критерію загальну суму квадратів відхилень результативної ознаки Y від її середнього значення розкладають на дві компоненти: суму квадратів відхилень, обумовлену регресією, і залишкову суму квадратів відхилень, обумовлену впливом інших факторів, не врахованих рівнянням регресії:

$$\sum(y_i-\overline{y})^2=\sum(y_{ip}-\overline{y})^2+\sum(y_i-y_{ip})^2 \qquad (10.23)$$

де y_{ip} – значення результативної ознаки, отримані з рівняння регресії; $\sum(y_{ip}-\overline{y})^2$ - сума квадратів відхилень, обумовлена регресією; $\sum(y_i-y_{ip})^2$ - залишкова сума квадратів відхилень.

Загальна сума квадратів відхилень результативної ознаки від свого середнього значення викликана впливом множини факторів. Якщо фактор X не має впливу на результативну ознаку Y, то вся його дисперсія обумовлена впливом інших факторів, і загальна сума квадратів відхилень збігається із залишковою. Якщо ж інші фактори не впливають на результативну ознаку Y, то залежність функціональна, і залишкова сума квадратів відхилень дорівнює нулю, а загальна сума квадратів відхилень збігається із сумою квадратів відхилень, обумовленою регресією.

Придатність лінії регресії для наступного прогнозу залежить від того, яка частина загальної варіації результативної ознаки Y обумовлена регресією. Очевидно, що коли сума квадратів відхилень, обумовлена регресією, набагато перевершує залишкову суму квадратів відхилень, то рівняння регресії статистично значущо, і фактор X впливає на результативну ознаку Y.

Сума квадратів відхилень завжди пов'язана з числом ступенів свободи незалежного варіювання результативної ознаки. Число ступенів свободи визначається кількістю одиниць вибірки n і числом обумовлених параметрів. Число ступенів свободи показує, скільки незалежних відхилень з n можливих необхідно для утворення даної суми квадратів. Зокрема, для загальної суми квадратів відхилень потрібно n-1 незалежних відхилень. Можна показати, що факторна сума квадратів відхилень при лінійній регресії залежить тільки від однієї постійної – коефіцієнта регресії ρ_{yx}, тобто має один ступінь свободи. Числа ступенів свободи загальної, факторної і залишкової сум квадратів відхилень пов'язані співвідношенням

$$n-1 = 1+(n-2).$$

Відношення загальної, факторної і залишкової сум квадратів відхилень до відповідного числа ступенів свободи дає дисперсію, що припадає на один ступінь свободи. Визначення дисперсії на один ступінь свободи приводить їх до порівнянного вигляду. Значення F-критерію обчислюють як відношення факторної і залишкової дисперсій:

$$F_{спост} = \frac{S_{факт}}{S_{ост}} \tag{10.24}$$

Якщо нульова гіпотеза справедлива, то факторна і залишкова дисперсії не відрізняються. Для спростування нульової гіпотези необхідно, щоб факторна дисперсія перевищувала залишкову в кілька разів. Статистичні таблиці містять значення F-критерію при різних рівнях істотності нульової гіпотези і різному числі ступенів свободи. Табличне значення F-критерію являє собою максимальне значення відношення факторної і залишкової дисперсій при випадковій їх розбіжності. Таким чином, якщо обчислене значення F-критерію перевищує табличне, то нульову гіпотезу треба відкинути. У противному разі, тобто якщо обчислене значення F-критерію менше табличного, нульову гіпотезу вважають справедливою, а рівняння регресії - статистично незначущим.

З метою оцінки значущості параметрів лінійної регресії ρ і b перевіряють нульову гіпотезу, використовуючи t-розподіл Стьюдента при числі ступенів свободи n-2. Для цього розраховують стандартні помилки відповідно m_ρ і m_b, а потім значення t-критерію.

Стандартна помилка параметрів лінійної регресії визначається за формулами

$$m_\rho = \sqrt{\frac{S_{ост}}{\sum(x_i - \bar{x})^2}}; \qquad m_b = \sqrt{\frac{S_{ост} * \sum x_i^2}{n * \sum(x_i - \bar{x})^2}}. \qquad (10.25)$$

Розрахункове значення t-критерію Стьюдента знаходимо як відношення відповідного коефіцієнта до його стандартної помилки:

$$t_\rho = \frac{\rho}{m_\rho}; \qquad t_b = \frac{b}{m_b} \qquad (10.26)$$

Якщо розрахункове значення t-критерію перевищує табличне, то гіпотезу про неістотність параметра можна відхилити.

Довірчий інтервал для параметрів лінійної регресії визначається в такий спосіб:

$$\rho \pm t_{табл} * m_\rho; \qquad b \pm t_{табл} * m_b. \qquad (10.27)$$

Для оцінки значущості коефіцієнта кореляції розраховують його помилку m_r і значення t-критерію за формулами

$$m_r = \sqrt{\frac{1 - r^2}{n - 2}}; \qquad t_r = \frac{r}{m_r}. \qquad (10.28)$$

Можна показати, що $t_r = t_\rho$, отже перевірка гіпотез про значущість цих коефіцієнтів рівнозначна перевірці гіпотези про істотність лінійного рівняння регресії. Якщо коефіцієнт кореляції близький до одиниці, вводять допоміжну величину z:

$$z = \frac{1}{2} \ln \frac{1 + r}{1 - r}.$$

Стандартну помилку величини z і значення t-критерію знаходимо за формулами

$$m_z = \frac{1}{\sqrt{n - 3}}; \qquad t_z = \frac{z}{m_z}.$$

Коефіцієнт кореляції значущо відрізняється від нуля, якщо розрахункове значення t-критерію перевищує його табличну величину.

Для оцінки якості підбору лінійної функції розраховують квадрат коефіцієнта кореляції, іменований коефіцієнтом детермінації r^2. Коефіцієнт детермінації r^2 характеризує частку дисперсії результативної ознаки y, що пояснюється регресією, у загальній дисперсії. Чим більше частка поясненої дисперсії, тим менше роль інших факторів у варіації результативної ознаки y, і тим краще лінійна модель апроксимує статистичні дані. Гарною лінійною моделлю можна користуватися для прогнозу значень результативної ознаки y, що лежать за межами досліджуваної вибірки.

Приклад.10.9. Перевіримо на адекватність рівняння лінійної регресії з прикладу 9.1.

$$Y = -1/3 + 1,5\ X$$

x_i	1	2	3
y_i	1	3	4
y_{ip}	1,17	2,67	4,17
$(y_{ip} - \overline{y})^2$	2,2400	0,0000	2,2600
$(y_i - y_{ip})^2$	0,0289	0,1089	0,0289

Розв'язання: Обчислимо факторну і залишкову дисперсії:

$$S_{факт} = \frac{2,24 + 0,0 + 2,26}{1} = 4,5; \qquad S_{ост} = \frac{0,0289 + 0,1089 + 0,0289}{3-2} = 0,1667.$$

Знайдемо розрахункове значення F-критерію:

$$F_{спост} = \frac{4,5}{0,1667} = 26,995.$$

Отримане значення $F_{спост}$=26,995 порівнюємо з табличним $F_т$. Табличне значення $F_т$ = 163 знаходимо залежно від числа ступенів свободи для факторної і залишкової дисперсій. Оскільки $F_{спост}$=26,995 < $F_т$ -= 163, нульова гіпотеза повинна бути прийнята. Таким чином, рівняння регресії не є статистично значущим.

Оцінку значущості рівняння регресії звичайно проводять у вигляді таблиці дисперсійного аналізу:

Джерело варіації	Число ступенів свободи	Сума квадратів відхилень	Дисперсія на один ступінь свободи	F-критерій	
				розрахунковий	табличний при α=0,05
Загальна	2	4,667			
Регресія	1	4,5	4,5	26,995	163
Остаток	1	0,1667	0,1667		

Приклад 10.10. Перевірити статистичну значущість рівняння лінійної регресії з прикладу 9.2:

$$y = 0,00796x - 0,0034.$$

Розв'язання. Для перевірки статистичної значущості рівняння лінійної регресії скористаємося критерієм Фішера, для чого визначимо загальну, факторну і залишкову суми квадратів відхилень

№ досліду	x_i	y_i	y_{ip}	$(y_{ip}-\bar{y})^2$	$(y_i-y_{ip})^2$	$(x_i-\bar{x})^2$	x_i^2
1	4	0,041	0,02844	0,018502	0,0001578	291,62	16
2	8	0,05	0,06028	0,010854	0,0001057	171,01	64
3	10	0,081	0,0762	0,007790	0,0000230	122,70	100
4	14	0,104	0,10804	0,003183	0,0000163	50,08	196
5	16	0,12	0,12396	0,001640	0,0000157	25,78	256
6	20	0,139	0,1558	0,000075	0,0002822	1,16	400
7	19	0,154	0,14784	0,000276	0,0000379	4,31	361
8	23	0,18	0,17968	0,000232	0,0000001	3,70	529
9	26	0,208	0,20356	0,001529	0,0000197	24,24	676
10	30	0,241	0,2354	0,005032	0,0000314	79,62	900
11	31	0,25	0,24336	0,006225	0,0000441	98,47	961
12	36	0,269	0,28316	0,014089	0,0002005	222,70	1296
13	37	0,301	0,29112	0,016042	0,0000976	253,54	1369
Сума	Σx_i=274	Σy_i=2,138		0,085470	0,001032	1348,92	7124
Середнє	21,08	0,16446					
Дисперсія	112,41	0,00722	0,00712				

Визначимо число ступенів свободи і знайдемо загальну, факторну та залишкову дисперсії. Для визначення F-критерію скористаємося формулою (10.21). Розрахунки зведемо в таблицю

Джерело варіації	Число ступенів свободи		Сума квадратів відхилень	Дисперсія на один ступінь свободи	F-критерій	
	формула	значення			розрахунковий	табличний при α=0,05
Загальна	n-1	12	0,0865	0,00721		
Регресія	m-1	1	0,0855	0,0855	910,977	4,84
Остача	n-m	11	0,00103	0,0000938		

Оскільки F = 910,98 > $F_т$ =4,84, рівняння лінійної регресії має статистичну значущість.

Для розрахунку довірчих інтервалів для параметрів лінійної регресії визначимо стандартні помилки m_ρ і m_b:

$$m_\rho = \sqrt{\frac{0{,}0000938}{1348{,}92}} = 0{,}00026373; \qquad m_b = \sqrt{\frac{0{,}0000938*7124}{13*1348{,}92}} = 0{,}00617376$$

і визначимо розрахункові значення t-критерію Стьюдента

$$t_\rho = \frac{0{,}00796}{0{,}00026373} = 30{,}1824; \qquad t_b = \frac{0{,}0034}{0{,}00617376} = 0{,}5507.$$

Табличне значення t-критерію t = 2,2. Розрахункове значення t-критерію для параметра ρ перевищує табличне $t_ρ$=30,1824 > $t_{ρт}$=2,2, а розрахункове значення t-критерію для параметра b менше табличного t_b=0,5507 < $t_{bт}$=2,2, тобто параметр ρ є значущим, а параметр b не значущий, його можна дорівняти нулю.

95-відсотковий довірчий інтервал для параметра ρ визначиться в такий спосіб: ρ±2,2*6,955*10^{-8}, маємо

$$0{,}00738 \le \rho \le 0{,}00854.$$

Для оцінки значущості коефіцієнта кореляції розрахуємо його помилку m_r і значення t-критерію за формулами (10.26):

$$m_r = \sqrt{\frac{1-0{,}994^2}{13-2}} = 0{,}003298; \qquad t_r = \frac{0{,}994}{0{,}003298} = 30{,}14 \ .$$

Табличне значення t-критерію дорівнює 2,2, тобто t_r = 30,14 >$t_{rт}$=2,2, отже лінійний зв'язок між результативною ознакою Y і фактором X істотний. Перевіримо висновок, використовуючи величину z, тому що коефіцієнт кореляції дуже близький до одиниці. Розрахуємо z, її стандартну помилку і значення t-критерію:

$$z = \frac{1}{2}\ln\frac{1+0{,}994}{1-0{,}994} = 2{,}903; \quad m_z = \frac{1}{\sqrt{13-3}} = 0{,}3162; \qquad t_z = \frac{2{,}903}{0{,}3162} = 9{,}18 \ .$$

Табличне значення t-критерію дорівнює 2,2, маємо t_z = 9,18 >$t_{zт}$=2,2. Таким чином, істотність лінійного зв'язку підтверджується.

Приклад 10.11. Перевірити статистичну значущість рівняння лінійної регресії з прикладу 9.3, використовуючи функції Microsoft Excel

$$u = 1{,}525i + 3{,}948.$$

Розв'язання. Скористаємося статистичною функцією Microsoft Excel ЛИНЕЙН, що повертає масив, який описує отриману пряму. Функція задається у вигляді формули масиву, її аргумент Статистика вказує, чи потрібно повернути додаткову статистику по регресії.

Оскільки функція повертає масив, виділимо в аркуші Excel дві групи по п'ять комірок, в яких розмістяться параметри в наступному порядку:

значення коефіцієнта регресії $ρ_{yx}$	значення параметру b
стандартна помилка коефіцієнта регресії $m_ρ$	стандартна помилка параметру m_b
коефіцієнт детермінації r^2	стандартна помилка для прогнозних значень y m_y
розрахункове значення F-критерію	число ступенів свободи
факторна сума квадратів відхилень $(y_{ip}-\bar{y})^2$	залишкова сума квадратів відхилень $(y_i - y_{ip})^2$

У діапазон E11:F15 введемо функцію ЛИНЕЙН. Як аргумент Известные_значения_y вкажемо діапазон C2:C21, як аргумент Известные_значения_x - діапазон B2:B21. Щоб дістати ненульове значення b, в аргумент Конст введемо значення ИСТИНА. Оскільки нам потрібні ще інші дані, в аргумент Статистика також введемо значення ИСТИНА. Натиснемо F2 і завершимо введення функції комбінацією клавіш Shift+Ctrl+Enter (рис. 10.7).

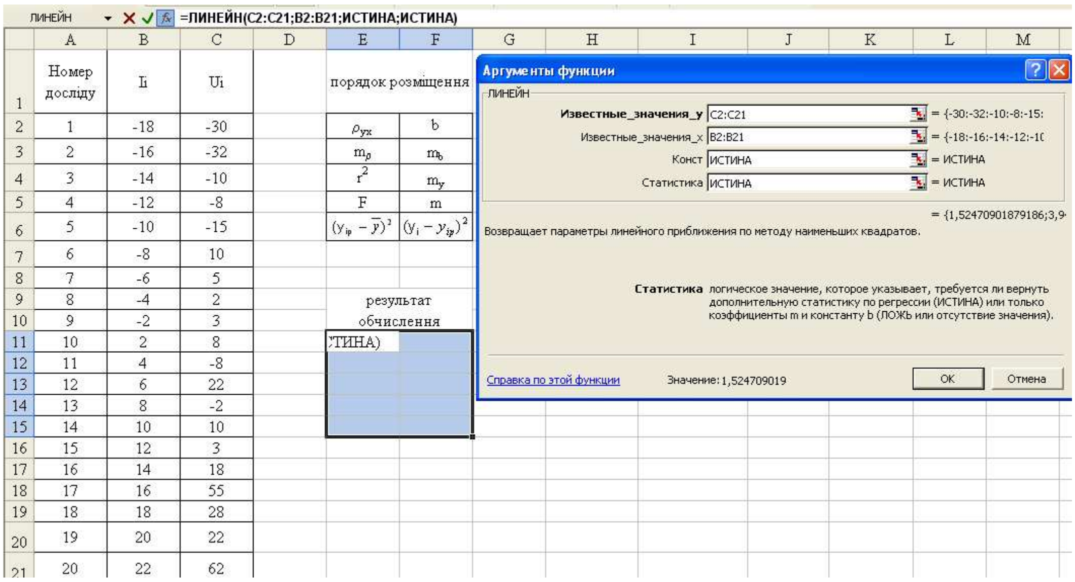
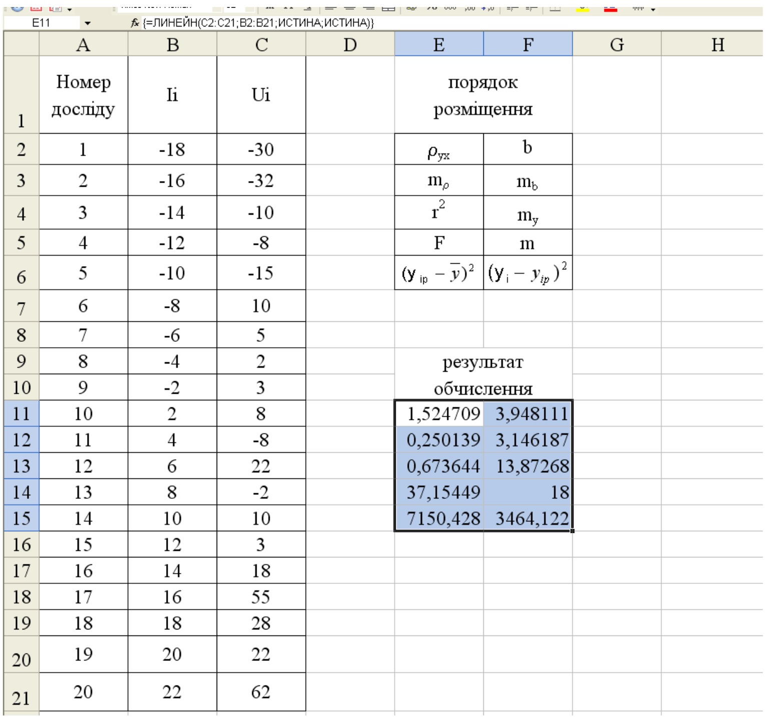


Рис. 10.7 - Введення аргументів функції ЛИНЕЙН

У результаті дістанемо



Номер досліду	Ii	Ui		порядок розміщення	
1	-18	-30		ρ_{yx}	b
2	-16	-32		m_ρ	m_b
3	-14	-10		r^2	m_y
4	-12	-8		F	m
5	-10	-15		$(y_{ip}-\bar{y})^2$	$(y_i-y_{ip})^2$
6	-8	10			
7	-6	5			
8	-4	2		результат обчислення	
9	-2	3			
10	2	8		1,524709	3,948111
11	4	-8		0,250139	3,146187
12	6	22		0,673644	13,87268
13	8	-2		37,15449	18
14	10	10		7150,428	3464,122
15	12	3			
16	14	18			
17	16	55			
18	18	28			
19	20	22			
20	22	62			

Рис. 10.8 - Результат розрахунку для перевірки статистичної значущості рівняння u = 1,525i + 3,948

Скористаємося результатами розрахунку для заповнення таблиці дисперсійного аналізу

Джерело варіації	Число ступенів свободи		Сума квадратів відхилень	Дисперсія на один ступінь свободи	F-критерій	
	формула	значення			розрахунковий	табличний при α=0,05
Загальна	n-1	19	10614,55	558,661		
Регресія	m-1	1	7150,428	7150,428	37,154	4,414
Залишок	n-m	18	3464,12	192,45		

Табличне значення F-критерію визначимо за допомогою статистичної функції FРАСПОБР, що повертає зворотне значення для F-розподілу імовірностей (рис. 10.9). Синтаксис:

FРАСПОБР(вероятность;степени_свободы1; степени_свободы2),

де Вероятность — це імовірність, пов'язана з F-розподілом; Степени_свободы1 — це чисельник ступенів свободи; Степени_свободы2 — це знаменник ступенів свободи.

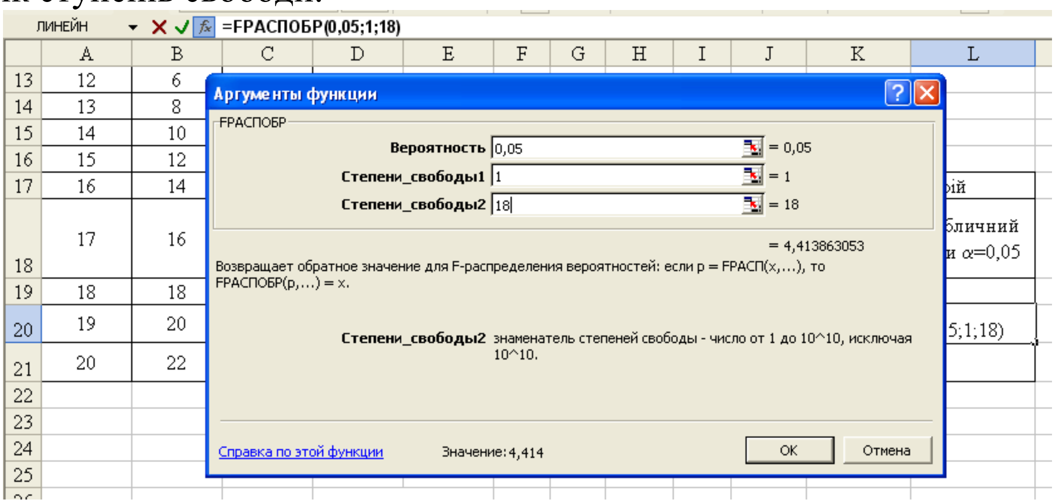


Рис. 10.9 - Введення аргументів функції FРАСПОБР

Оскільки F = 37,154 > $F_т$ =4,41, рівняння лінійної регресії має статистичну значущість.

Для розрахунку довірчих інтервалів для параметрів лінійної регресії скористаємося стандартними помилками m_ρ = 0,250139 і m_b = 3,146 і визначимо розрахункові значення t-критерію Стьюдента:

$$t_\rho = \frac{1{,}524709}{0{,}250139} = 6{,}095\,; \qquad t_b = \frac{3{,}948}{3{,}146} = 1{,}255\,.$$

Для знаходження табличного значення t-критерію скористаємося статистичною функцією СТЬЮДРАСПОБР, що повертає t-значення розподілу Стьюдента як функцію імовірності і числа ступенів свободи (рис. 10.10). Синтаксис

СТЬЮДРАСПОБР(вероятность;степени_свободы),

де Вероятность — імовірність, що відповідає двосторонньому розподілу Стьюдента; Степени_свободы — число ступенів свободи, що характеризує розподіл.

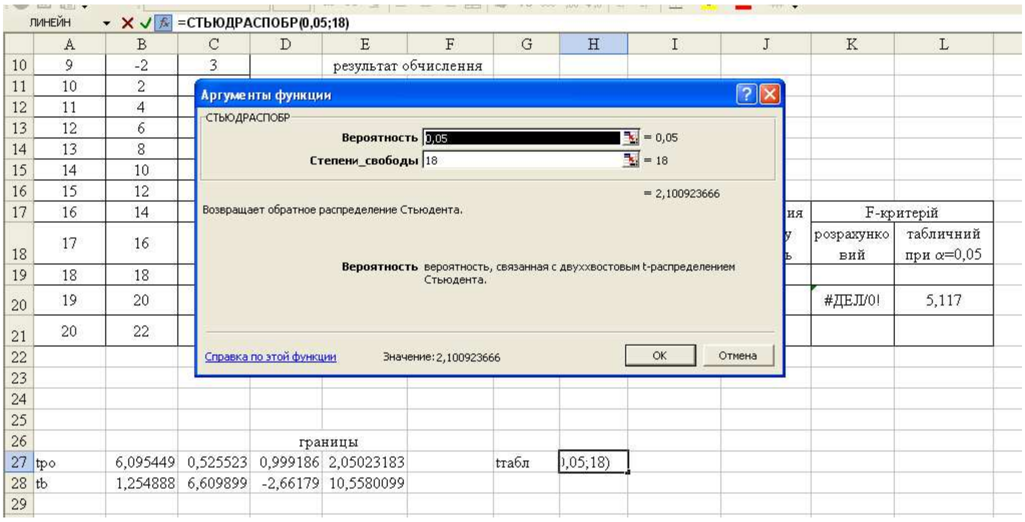


Рис. 10.10 - Введення аргументів функції СТЬЮДРАСПОБР

Табличне значення t-критерію t = 2,1. Розрахункове значення t-критерію для параметра ρ перевищує табличне t_ρ=6,095 > $t_{\rho т}$=2,1, а розрахункове значення t-критерію для параметра b менше табличного t_b=1,255 < $t_{bт}$=2,1, тобто параметр ρ є значущим, а параметр b не значущий, його можна дорівняти нулю.

95-відсотковий довірчий інтервал для параметра ρ визначиться в такий спосіб: ρ±2,1*0,250 і дістанемо

$$0{,}999 \le \rho \le 2{,}05.$$

Значення коефіцієнта детермінації r^2= 0,67 говорить про ступінь тісноти лінійного зв'язку. Для оцінки значущості коефіцієнта кореляції r= √0,67 = 0,82 розрахуємо його помилку m_r і значення t-критерію за формулами (10.26)

$$m_r = \sqrt{\frac{1-0{,}67}{20-2}} = 0{,}135; \qquad t_r = \frac{0{,}82}{0{,}135} = 6{,}095 .$$

Табличне значення t-критерію дорівнює 2,1, тобто t_r = 6,095 >$t_{rт}$=2,1, отже лінійний зв'язок між результативною ознакою U і фактором I істотний. Перевіримо висновок, використовуючи величину z. Визначимо z за допомогою статистичної функції ФИШЕР, що повертає перетворення Фішера для аргументу x. Це перетворення будує функцію, яка має нормальний, а не асиметричний розподіл (рис. 10.11). Синтаксис:

ФИШЕР(x),

де x — числове значення, яке потрібно перетворити.

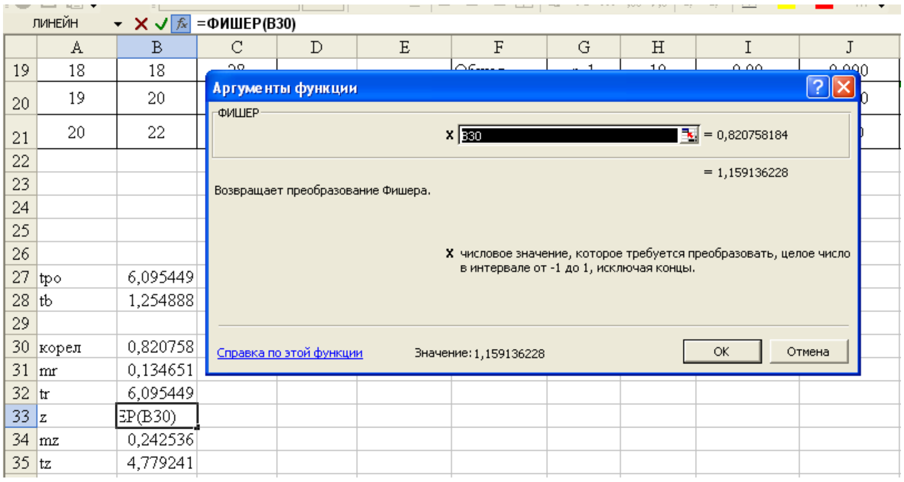


Рис. 10.11 - Введення аргументу функції ФИШЕР

Розрахуємо стандартну помилку z і значення t-критерію:

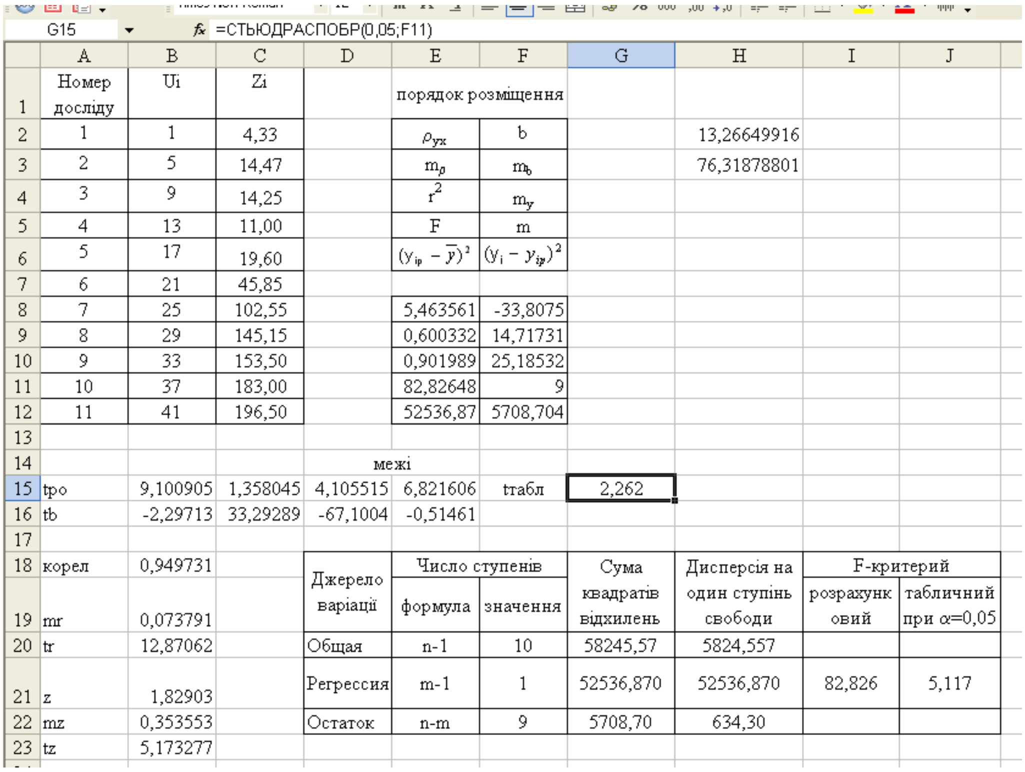
$$m_z = \frac{1}{\sqrt{20-3}} = 0{,}2425\,; \qquad t_z = \frac{1{,}159}{0{,}2425} = 4{,}779\ .$$

Табличне значення t-критерію дорівнює 2,1, маємо $t_z = 4{,}779 > t_{zт} = 2{,}1$. Таким чином, істотність лінійного зв'язку підтверджується.

Приклад 10.12. Перевірити статистичну значущість рівняння лінійної регресії з прикладу 9.4

$$z = 5{,}46u - 33{,}8.$$

Розв'язання. Скористаємося статистичними функціями Microsoft Excel (рис. 10.12)



G15 =СТЬЮДРАСПОБР(0,05;F11)

	A	B	C	D	E	F	G	H	I	J
1	Номер досліду	Ui	Zi		порядок розміщення					
2	1	1	4,33		ρ_{yx}	b		13,26649916		
3	2	5	14,47		m_ρ	m_b		76,31878801		
4	3	9	14,25		r^2	m_y				
5	4	13	11,00		F	m				
6	5	17	19,60		$(y_{ip}-\bar{y})^2$	$(y_i-y_{ip})^2$				
7	6	21	45,85							
8	7	25	102,55		5,463561	-33,8075				
9	8	29	145,15		0,600332	14,71731				
10	9	33	153,50		0,901989	25,18532				
11	10	37	183,00		82,82648	9				
12	11	41	196,50		52536,87	5708,704				
13										
14				межі						
15	tpo	9,100905	1,358045	4,105515	6,821606	tтабл	2,262			
16	tb	-2,29713	33,29289	-67,1004	-0,51461					
17										
18	корел	0,949731		Джерело варіації	Число ступенів		Сума квадратів відхилень	Дисперсія на один ступінь свободи	F-критерий	
19	mr	0,073791			формула	значення			розрахунковий	табличний при α=0,05
20	tr	12,87062		Общая	n-1	10	58245,57	5824,557		
21	z	1,82903		Регрессия	m-1	1	52536,870	52536,870	82,826	5,117
22	mz	0,353553		Остаток	n-m	9	5708,70	634,30		
23	tz	5,173277								

Рис. 10.12 - Визначення статистичної значущості рівняння лінійної регресії $z = 5{,}46u - 33{,}8$

Заповнимо таблицю для дисперсійного аналізу

Джерело варіації	Число ступенів свободи		Сума квадратів відхилень	Дисперсія на один ступінь свободи	F-критерій	
	формула	значення			розрахунковий	табличний при α=0,05
Загальна	n-1	10	58245,57	5824,557		
Регресія	m-1	1	52536,87	52536,87	82,826	5,117
Остаток	n-m	9	5708,7	634,3		

Оскільки F = 82,826 > $F_т$ =5,117, рівняння лінійної регресії статистично незначуще.

Розрахуємо довірчі інтервали для параметрів лінійної регресії. Скористаємося для цього стандартними помилками m_ρ = 0,6 і m_b = 14,7 і визначимо розрахункові значення t-критерію Стьюдента

$$t_\rho = \frac{5,46}{0,6} = 9,1; \qquad t_b = \frac{33,81}{14,7} = 2,297.$$

Розрахункове значення t-критерію для параметра ρ перевищує табличне t_ρ=9,1 > $t_{\rho т}$=2,26, розрахункове значення t-критерію для параметра b також більше табличного t_b=2,297 > $t_{bт}$=2,26, тобто параметри лінійної регресії ρ і b є статистично значущими.

95-відсотковий довірчий інтервал для параметрів ρ і b визначиться в такий спосіб: ρ±2,26*0,6; b±2,26*14,7. Одержимо

$$4,1 \le \rho \le 6,82; \qquad -67,1 \le b \le -0,5.$$

Значення коефіцієнта детермінації r^2= 0,9 говорить про сильний лінійний зв'язок. Оцінимо значущість коефіцієнта кореляції r= √0,9 = 0,95. Розрахуємо помилку m_r і значення t-критерію за формулами (10.26):

$$m_r = \sqrt{\frac{1-0,902}{11-2}} = 0,074; \qquad t_r = \frac{0,95}{0,074} = 12,87 .$$

Табличне значення t-критерію дорівнює 2,26, тобто t_r = 12,87 > $t_{rт}$=2,26, отже лінійний зв'язок між результативною ознакою Z і фактором U статистично значущий.

Застосуємо величину z для перевірки нульової гіпотези, тому що коефіцієнт кореляції близький до одиниці і число дослідів невелике. На підставі отриманого значення z=1,83 розрахуємо її стандартну помилку і значення t-критерію:

$$m_z = \frac{1}{\sqrt{11-3}} = 0,354; \qquad t_z = \frac{1,83}{0,354} = 5,173 .$$

Табличне значення t-критерію дорівнює 2,26, маємо t_z = 5,173 >$t_{zт}$=2,26. Таким чином, істотність лінійного зв'язку підтверджується.

Оскільки рівняння лінійної регресії виявилося статистично незначущим за критерієм Фішера, а його параметри незначущими не визнаються, слід перевірити гіпотезу про квадратичну залежність між фактором U і результативною ознакою Z. Гіпотеза про квадратичну залежність між фактором U і результативною ознакою Z більш імовірна.

Запитання для самоперевірки:

1. Поясніть, що таке статистична гіпотеза?
2. Поясніть, що таке нульова і альтернативна гіпотези?
3. Які критерії застосовують для перевірки статистичних гіпотез?
4. Поясніть значення термінів «критична область», «критична точка», «область прийняття гіпотези».
5. В якому випадку слід вибирати двосторонню критичну область?
6. Який результат перевірки гіпотези відносять до помилки 1-го роду і який до помилки 2-го роду?
7. Наведіть приклади задач на перевірку гіпотез.
8. Які задачі вирішують за допомогою однофакторного дисперсійного аналізу?
9. Чому цей вид аналізу випадкових даних отримав назву дисперсійного?
10. Що являють собою величини $S_{общ}$, $S_{факт}$ і $S_{ост}$? Яке співвідношення між ними?
11. Які задачі вирішують методом регресійного аналізу?
12. Для чого застосовують метод найменших квадратів?
13. Що таке рівняння регресії? Як перевіряють його адекватність статистичним даним?
14. Як визначають значущість коефіцієнтів рівняння регресії?

Задачі для самостійного розв'язання

10.1. Число появ герба при 20 підкиданнях двох монет розподілилося в такий спосіб:

Число гербів	0	1	2
Число появ	4	8	8

Чи узгоджуються ці дані з припущенням про симетричність монет і незалежність результатів підкидання. Як рівень значущості прийняти α=0,05.

Змістовий модуль 3. ВИПАДКОВІ ПРОЦЕСИ

Тема 11. ЕЛЕМЕНТИ ТЕОРІЇ ВИПАДКОВИХ ПРОЦЕСІВ

Випадковий процес

Функція X аргументу t називається випадковою, якщо при кожному заданому значенні аргументу t величина X є випадковою. Вигляд, прийнятий випадковою функцією X в результаті досліду, називається **реалізацією** функції X.

Процеси, описувані випадковими функціями, називаються **випадковими** або **стохастичними**.

На рис. 11.1 показане сімейство реалізацій випадкової функції X(t).

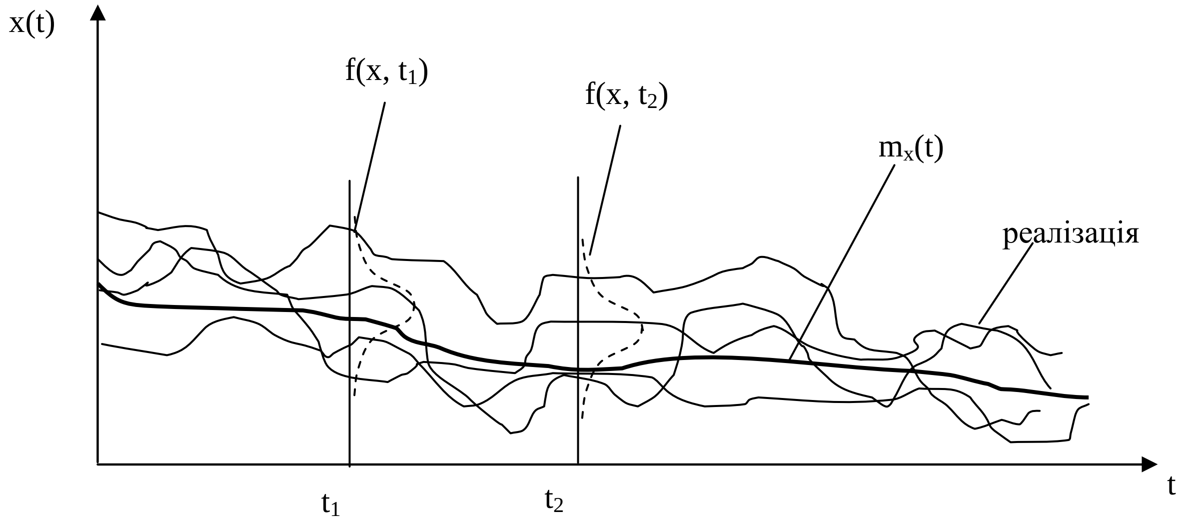


Рис. 11.1 - Сімейство реалізацій випадкової функції X(t)

Імовірнісні характеристики випадкового процесу є функціями часу. Якщо зафіксувати час t, то випадкова функція перетворюється у звичайну випадкову величину, що може приймати різні значення і має деякий закон розподілу зі своїми параметрами. Будемо називати цю величину перетином випадкової функції, що відповідає даному t.

Закон розподілу однієї випадкової величини є функція одного аргументу. Закон розподілу системи двох випадкових величин - функція двох аргументів і т.д. Випадкова функція X(t) називається нормальною, якщо закон розподілу системи будь-якого числа n її перетинів є n-мірним нормальним законом розподілу. Однак користування функціями багатьох аргументів як імовірнісними характеристиками настільки незручне, що звичайно розглядають тільки їх числові характеристики. Обмежимося розглядом найпростіших характеристик випадкових функцій, аналогічних числовим характеристикам випадкових величин. Апарат числових характеристик дозволяє порівняно просто вирішувати багато практичних задач.

На відміну від числових характеристик випадкових величин, що представляють собою певні числа, характеристики випадкових функцій являють собою не числа, а функції.

Математичним сподіванням випадкової функції X(t) називається невипадкова функція $m_x(t)$, яка при кожному значенні аргументу t дорівнює математичному сподіванню відповідного перетину випадкової функції:

$$m_x(t) = M[X(t)]. \tag{11.1}$$

Дисперсією випадкової функції X(t) називається невипадкова функція $D_x(t)$, значення якої для кожного t дорівнює дисперсії відповідного перетину випадкової функції.

$$D_x(t) = D[X(t)] \tag{11.2}$$

Таким чином, математичне сподівання - це деяка середня функція аргументу t, біля якої різним чином варіюються конкретні реалізації випадкової функції. Дисперсія випадкової функції при кожному t характеризує розкид можливих реалізацій випадкової функції відносно середньої.

Однак для опису основних особливостей випадкових функцій цих характеристик недостатньо.

Розглянемо сімейство реалізацій випадкових функцій $X_1(t)$ і $X_2(t)$ на рис. 11.2 і рис. 11.3.

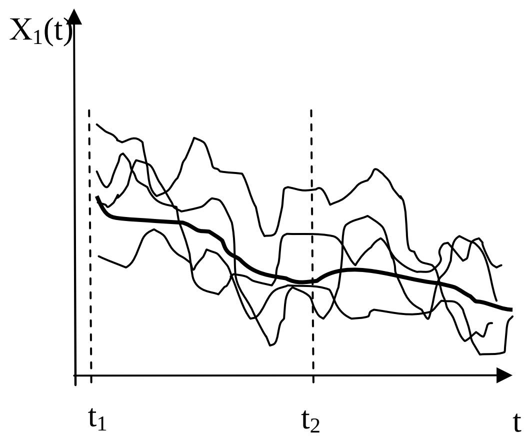


Рис. 11.2.

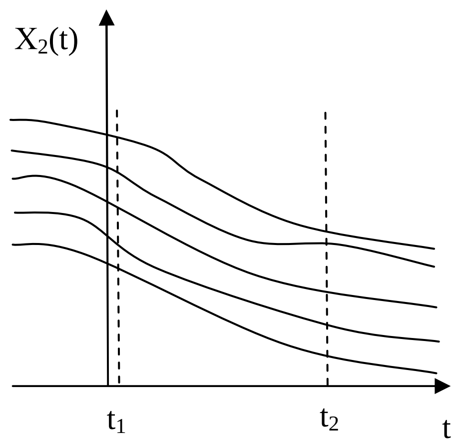


Рис. 11.3

У випадкових функцій $X_1(t)$ і $X_2(t)$ на рис. 11.2 і 11.3 приблизно однакові математичні сподівання і дисперсії, але характер цих випадкових функцій різко відрізняється. Для $X_2(t)$ характерна яскраво виражена залежність між її значеннями при різних t (якщо при t_1 вона прийняла значення менше середнього, то і при t_2 також менше середнього). Випадкова функція $X_1(t)$ має різко коливальний характер з безперервними безладними коливаннями. Для такої функції характерно швидке загасання залежності між її значеннями в міру збільшення відстані за t між ними.

Внутрішня структура розглянутих процесів зовсім різна, і для її опису вводять спеціальну характеристику - кореляційну функцію. Кореляційна функція характеризує ступінь залежності між перетинами випадкової функції, що належать до різних t.

Кореляційною функцією випадкової функції X(t) називається невипадкова функція двох аргументів K(t, t'), яка при кожній парі значень t, t' дорівнює кореляційному моменту відповідних перетинів випадкової функції:

$$K_x(t,t') = M[\overset{0}{X}(t) * \overset{0}{X}(t')] \tag{11.3}$$

Таким чином, якщо повернутися до рис. 11.2 і 11.3, кореляційна функція $X_2(t)$ повільно убуває зі збільшенням проміжку (t, t’), а кореляційна функція $X_1(t)$ убуває швидко.

Якщо аргументи кореляційної функції K_x збігаються, тобто t = t’, то

$$K_x(t) = M[\overset{0}{X}(t) * \overset{0}{X}(t)] = M[\overset{0}{X}{}^2(t)] = D_x(t), \tag{11.4}$$

тобто при t = t' кореляційна функція обертається в дисперсію випадкової функції X(t).

Таким чином, необхідність у визначенні дисперсії випадкової функції відпадає. Як основні характеристики випадкової функції достатньо розглядати її математичне сподівання і кореляційну функцію.

Кореляційна функція має такі основні властивості.

1. Є симетричною функцією відносно своїх аргументів

$$K_x(t,t') = K_x(t',t), \tag{11.5}$$

тобто не змінюється при заміні t на t'.

2. Кореляційна функція ніколи не перевищує добутки середніх квадратичних відхилень її аргументів:

$$|K_x(t,t')| \le \sigma_x(t)\sigma_x(t'). \tag{11.6}$$

3. Кореляційна функція K(t, t') - додатньо визначена, тобто її інтеграл для будь-якої області інтегрування не буває від’ємний:

$$\iint K_x(t,t')\varphi(t)\varphi(t')dtdt' \ge 0, \tag{11.7}$$

де φ(t) - будь-яка функція.

Для нормальної випадкової функції X(t) характеристики $m_x(t)$ і K(t, t’) є вичерпними і визначають закон розподілу будь-якого числа її перетинів.

При додаванні до випадкової функції X(t) невипадкового доданку φ(t) до її математичного сподівання додається той же невипадковий доданок, а кореляційна функція не змінюється (рис. 11.4).

Можна показати, що при множенні випадкової функції X(t) на невипадковий множник φ(t) її математичне сподівання множиться на той же множник φ(t), а кореляційна функція множиться на множник φ(t)φ(t').

Якщо випадкову функцію X(t) піддати деякому перетворенню A_t, то вийде інша випадкова функція Y(t):

$$Y(t) = A_t \{X(t)\}.$$

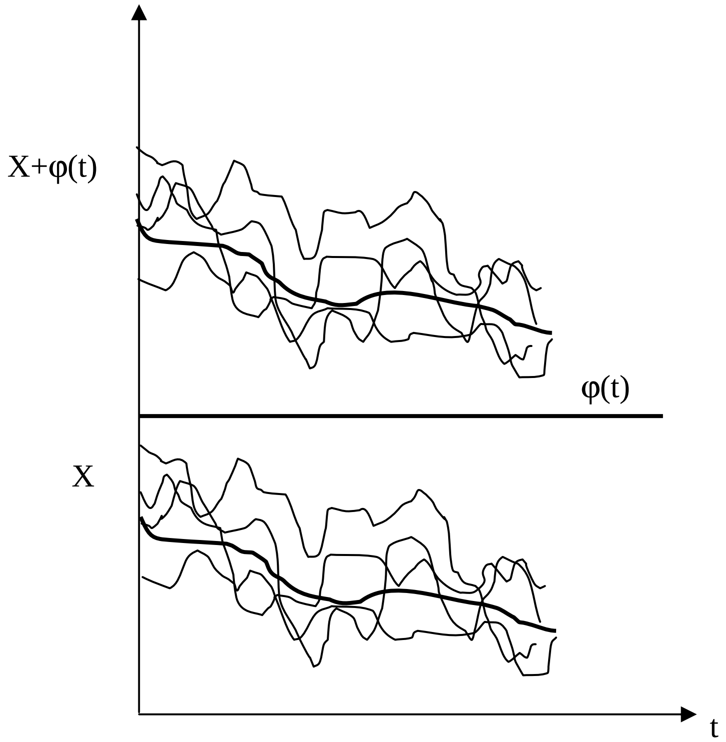


Рис. 11.4 - Додавання до випадкової функції X(t) невипадкового доданку φ(t)

Відомо, що перетворення називається лінійним однорідним, якщо його можна застосовувати до суми елементів почленно і при цьому постійний множник можна виносити за знак перетворення. Лінійне перетворення за аргументом t позначається L_t.

Лінійне перетворення суми випадкових функцій $X_i(t)$ дорівнює сумі перетворень від кожного доданка:

$$L_t\left\{\sum_{i=1}^{n} c_i X_i(t)\right\} = \sum_{i=1}^{n} c_i L_t\{X_i(t)\}, \tag{11.8}$$

де c_i - множник, що не залежить від аргументу t, за яким провадиться перетворення.

Якщо випадкова функція Y(t) зв'язана з випадковою функцією X(t) лінійним перетворенням

$$Y(t) = L_t\{X_i(t)\},$$

то її математичне сподівання $m_y(t)$ утворюється шляхом того ж лінійного перетворення $m_x(t)$

$$m_y(t) = L_t\{m_x(t)\}, \tag{11.9}$$

а для знаходження кореляційної функції $K_y(t, t')$ потрібно двічі піддати кореляційну функцію $K_x(t, t')$ відповідному лінійному перетворенню, один раз за t, другий раз за t':

$$K_y(t,t') = L_t\left\{L_{t'}\left\{K_x(t,t')\right\}\right\} \tag{11.10}$$

Стаціонарний випадковий процес

На практиці часто зустрічаються процеси, що протікають у часі приблизно однорідно і мають вигляд безперервних випадкових коливань навколо деякого середнього значення. Причому ні середня амплітуда, ні характер цих коливань не виявляють істотних змін з часом. Такі випадкові процеси називаються **стаціонарними**. Нестаціонарний же процес характерний тим, що він має тенденцію розвитку в часі.

Випадковий процес x(t) називається стаціонарним, якщо його імовірнісні характеристики не залежать від початку відліку часу, тобто закон розподілу перетинів той самий:

$$f(x, t_1) = f(x, t_2) = f(x).$$

Не залежать від початку відліку часу і числові характеристики - математичне сподівання і дисперсія:

$$m(t) = const;\ D_X(t) = const. \qquad (11.11)$$

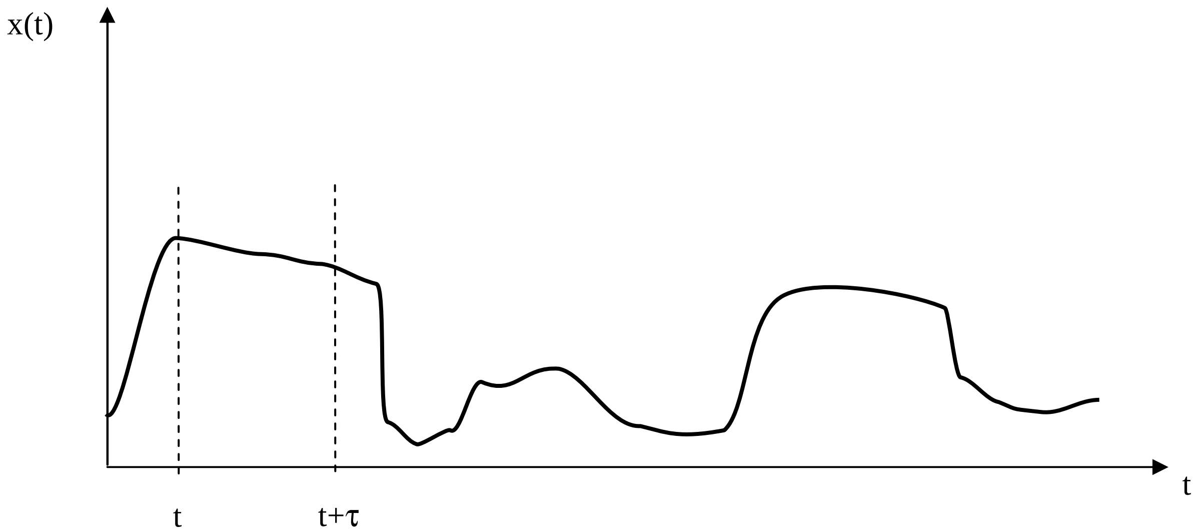


Рис. 11.5 - Перетини випадкової функції, що відповідають моментам часу t і t+τ

Установимо, якій умові повинна задовольняти кореляційна функція стаціонарного випадкового процесу. Покладемо $t'=t+\tau$ і розглянемо кореляційний момент двох перетинів x(t), розділений інтервалом часу τ, K(t, t+τ).
Очевидно, що цей кореляційний момент не повинен залежати від того, де саме на осі взятий інтервал τ, а повинен залежати тільки від довжини цього інтервалу, тобто

$$K(t_1, t_1+\tau) = K(\tau) = R_X(\tau). \qquad (11.12)$$

Таким чином, вираз (11.12) – єдина істотна умова, якій повинен задовольняти стаціонарний випадковий процес. Ця залежність називається кореляційною функцією $R_X(\tau)$. Якщо $\tau=0$, то кореляційна функція дорівнює дисперсії і максимальна:

$$R_X(\tau)_{/\tau=0} = D_X. \qquad (11.13)$$

Зі збільшенням τ зв'язок між перетинами стає слабкішим, тобто $R_X(0) \geq R_X(\tau)$, що показано на рис. 11.6.

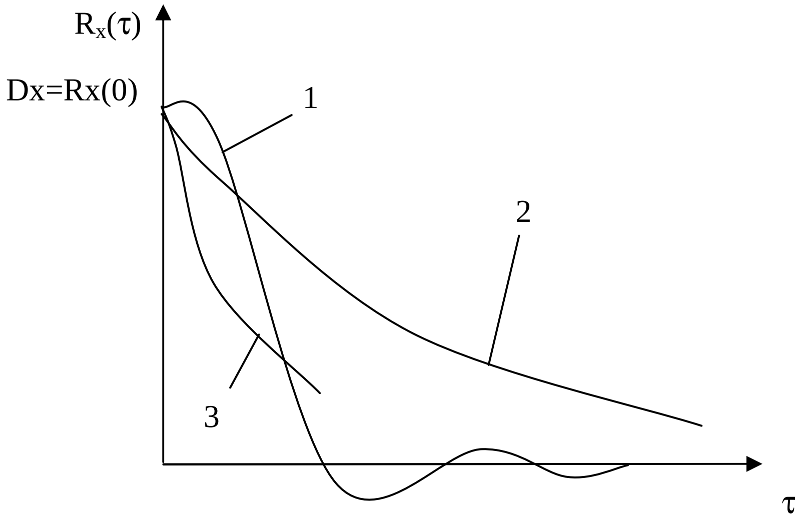


Рис. 11.6 - Криві кореляційних функцій

Із симетрії кореляційної функції K(t, t') випливає, що $R_X(\tau)=R_X(-\tau)$, тобто кореляційна функція стаціонарної випадкової функції є парна функція аргументу τ, її крива симетрична відносно осі ординат. Кореляційна функція характеризує внутрішній частотний склад випадкового процесу. Чим більше високочастотних складових містить випадковий процес, тим різкіше спадає зв'язок зі збільшенням τ (криві 2 і 3 на рис. 11.6). Коливання кореляційної функції (крива 1) вказують на приховану періодичність процесу.

Ергодична гіпотеза

Випадковий процес називається ергодичним, якщо всі його статистичні характеристики можуть бути визначені за однією реалізацією достатньої тривалості. Стаціонарний випадковий процес звичайно має властивість ергодичності. При цьому з імовірністю, рівною одиниці, числові характеристики, визначені за множиною реалізацій,

математичне сподівання $$M[x(t)] = \int_{-\infty}^{\infty} x(t)\, f[x(t)]dx \qquad (11.14)$$

і кореляційна функція (як другий змішаний центральний момент)

$$R_x(\tau) = M[\overset{0}{x}(t)\,\overset{0}{x}(t+\tau)] = \iint \overset{0}{x}(t)\,\overset{0}{x}(t+\tau)\, f(x_1, x_2)dx_1 dx_2 \qquad (11.15)$$

дорівнюють відповідним числовим характеристикам, визначеним за часом:

математичному сподіванню $$M[x(t)] = \lim \frac{1}{2T}\int_{-T}^{T} x(t)dt \qquad (11.16)$$

і кореляційної функції $$R_x(t) = \lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{T} \overset{0}{x}(t)\,\overset{0}{x}(t+\tau)dt\,. \qquad (11.17)$$

Достатньою умовою ергодичності стаціонарної випадкової функції є прагнення до нуля її кореляційної функції при $\tau \to \infty$:

$$\lim_{\tau \to \infty} R_x(\tau) = 0 . \qquad (11.18)$$

Обчислення характеристик стаціонарного ергодичного випадкового процесу за однією реалізацією проводять в наступній послідовності:

1. Для обчислення інтегралів (11.16) і (11.17) дискретизують задачу, тобто спостережуваний період часу розбивають на інтервали Δ. Інтервал дискретності вибирають на підставі теореми Котельникова:

$$\Delta \le \frac{\pi}{\omega_{max}},$$

де ω_{max} – максимальна частота, що міститься в реалізації.

При цьому кількість отриманих ординат буде N+1, де N=T/Δ.

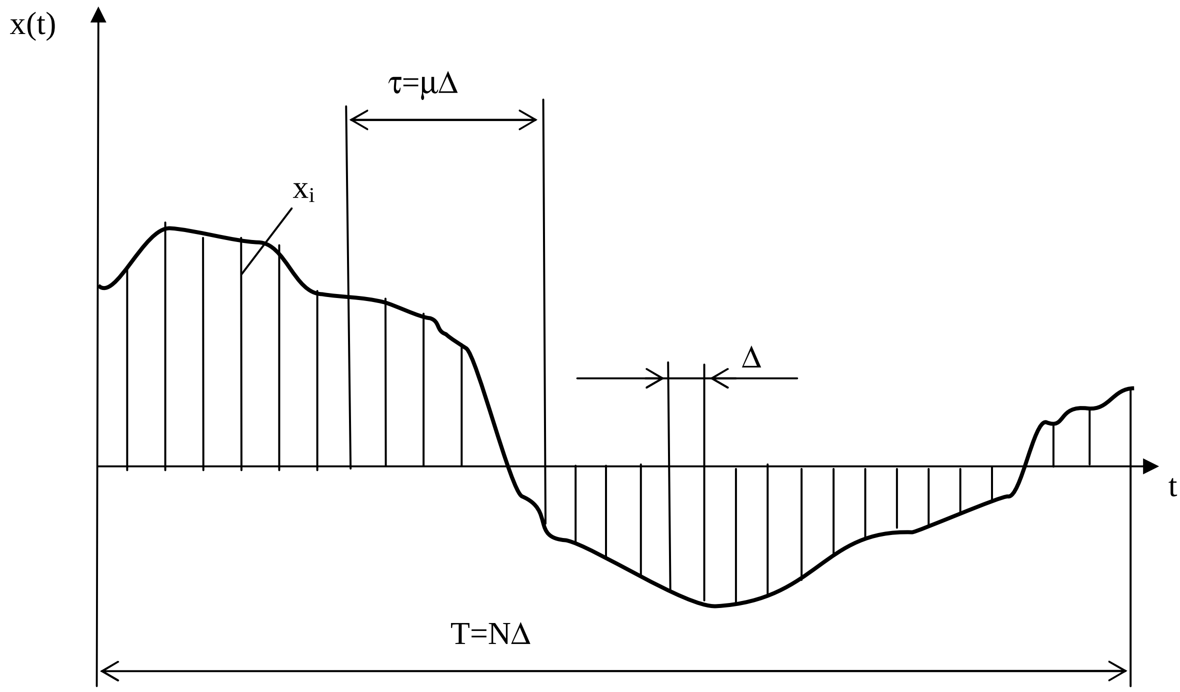


Рис. 11.7 - Обробка однієї реалізації стаціонарного ергодичного випадкового процесу

2. Обчислюють значення математичного сподівання, замінивши інтеграл (11.16) сумою

$$m_x = \frac{1}{N+1}\sum_{i=0}^{N} x_i \qquad (11.19)$$

3. Розраховують значення кореляційної функції, користуючись дискретним співвідношенням

$$R_x(\mu) = \frac{1}{N-\mu}\sum_{i=0}^{N-\mu} \overset{0}{x}_i \, \overset{0}{x}_{i+\mu} , \qquad (11.20)$$

де $\overset{0}{x}_i$ - центроване значення ознаки в момент часу t_i = $i*\Delta$, рівне $\overset{0}{x}_i = x_i - m_x$, $i = \overline{0, N-\mu}$; $\overset{0}{x}_{i+\mu}$ - центроване значення ознаки для моменту часу $t_{i+\mu}$.

Параметр μ визначають за формулою $\mu = \tau/\Delta = \overline{0, \mu_{\max}}$. Максимальне значення μ звичайно вибирають, виходячи з нижчої частоти, що міститься в реалізації:

$$\mu_{\max} = \frac{2\pi}{\omega_{низ} * \Delta} = \frac{\tau_{\max}}{\Delta}$$

Для одержання достовірних зведень про кореляційну функцію період спостереження Т повинен набагато перевищувати $\tau_{\max}$:

$$\tau_{\max} \le (0{,}1\text{-}0{,}2)\ T \quad \text{або} \quad \mu_{\max} \le (0{,}1\text{-}0{,}2)\ N.$$

4. Отримані експериментальні точки згладжують за методом найменших квадратів залежністю вигляду

$$R_x(\tau) = Ae^{-\alpha t}\cos\beta\tau, \tag{11.21}$$

де визначають параметри α і β, а параметр А знаходять із (11.20) при μ=0.

Приклад 11.1. Випадкова функція X(t) має характеристики $m_x(t)=1$ і $K_x(t,t') = e^{\alpha(t+t')}$. а) визначити характеристики випадкової функції

$$Y(t) = t\frac{dX(t)}{dt} + 1;$$

б) визначити, чи є стаціонарними випадкові функції X(t) і Y(t).

Розв'язання. а) Скористаємося тим, що перетворення $t\frac{dX(t)}{dt} + 1$ є лінійним, і за формулами (11.9) і (11.10) визначимо математичне сподівання і кореляційну функцію випадкової функції Y(t):

$$m_y(t) = t\frac{m_x(t)}{dt} + 1 = 1;$$

$$K_y(t,t') = L_t\{L_{t'}\{K_x(t,t')\}\} = L_t\left\{t\frac{dK(t,t')}{dt} + 1\right\} =$$

$$= t't\frac{d^2K(t,t')}{dtdt'} + 0 = t't\frac{d^2e^{\alpha(t+t')}}{dtdt'} = t't\alpha^2e^{\alpha(t+t')};$$

б) математичні сподівання випадкових функцій X(t) і Y(t) не залежать від аргументу t, а от їхні кореляційні функції залежать не тільки від відстані між перетинами τ=t' - t, але й від кожного з аргументів t і t'. Тому функції X(t) і Y(t) не є стаціонарними.

Приклад 11.2. Випадкова функція X(t) має характеристики

$$m_x(t)=0;\ K_x(t,t') = \frac{1}{1+(t'-t)^2}.$$

а) визначити характеристики випадкової функції

$$Y(t)=\int_0^t X(t)dt.$$

б) визначити, чи є стаціонарними випадкові функції X(t) і Y(t).

Розв'язання. а) визначимо математичне сподівання і кореляційну функцію випадкової функції Y(t) за формулами (11.9) і (11.10):

$$m_y(t)=\int_0^t m_x(t)dt=0;$$

$$K_y(t,t')=L_t\{L_{t'}\{K_x(t,t')\}\}=L_t\left\{\int_0^{t'}\frac{1}{1+(t'-t)^2}dt'\right\}=$$

$$=\int_0^t\left(\int_0^{t'}\frac{1}{1+(t'-t)^2}dt'\right)dt=\int_0^t\left(arctg(t'-t)-arctg(-t)\right)dt=$$

$$=tarctgt+t'arctgt'-(t-t')arctg(t-t')-0{,}5\ln\frac{(1+t^2)(1+t'^2)}{1+(t-t')^2};$$

б) кореляційна функція випадкової функції X(t) залежить тільки від відстані між перетинами (t'-t), отже X(t) є стаціонарною. Визначимо дисперсію випадкової функції Y(t)

$$D_y(t)=K_y(t,t)=2tarctgt-0{,}5\ln\frac{(1+t^2)^2}{1}=2tarctgt-\ln(1+t^2).$$

Дисперсія Y(t) залежить від аргументу t, отже випадкова функція Y(t) не є стаціонарною.

Приклад 11.3. Нехай є сума двох некорельованих випадкових функцій X(t) і Y(t) з характеристиками

$$m_x(t)=t;\ K_x(t,t')=tt';$$

$$m_y(t)=-t;\ K_y(t,t')=tt'e^{\alpha(t+t')}.$$

Знайти математичне сподівання і кореляційну функцію цієї суми.

Розв'язання. Позначимо суму випадкових функцій Z(t), тоді

$$Z(t) = X(t) + Y(t).$$

Визначимо математичне сподівання випадкової функції Z(t):

$$m_z(t) = m_x(t) + m_y(t) = t+(-t) = 0.$$

Визначимо кореляційну функцію випадкової функції Z(t):

$$K_z(t,t')=K_z(t,t')+K_y(t,t')=tt'+tt'e^{\alpha(t+t')}=tt'(1+e^{\alpha(t+t')}).$$

Елементи теорії масового обслуговування.
Основні поняття. Класифікація СМО

При дослідженні складних систем доводиться зустрічатися з системами, які призначені для багаторазового використання при вирішенні однотипних задач. Виникаючі при цьому процеси отримали назву процесів обслуговування, а системи - систем масового обслуговування (СМО). Системи масового обслуговування - це такі системи, в які у випадкові моменти часу надходять **заявки на обслуговування**, при цьому заявки, що надійшли, обслуговуються за допомогою наявних у розпорядженні системи **каналів** обслуговування. Обслуговування заявки відбувається протягом деякого часу $t_{об}$, величина якого випадкова. Потім канал звільняється і готовий до прийому наступної заявки. Випадковий характер потоку заявок і часу обслуговування $t_{об}$ приводить до того, що в якісь періоди часу на вході СМО нагромаджується надто велика кількість заявок. Вони або стають у чергу, або залишають СМО не обслуженими. В інші ж періоди часу СМО працюватиме з недовантаженням або простоювати. Стан СМО змінюється стрибком у моменти появи нової заявки або закінчення обслуговування, або в момент, коли заявка залишає чергу.

Прикладами систем масового обслуговування можуть служити станції технічного обслуговування автомобілів; персональні комп'ютери, що виконують вирішення тих або інших задач; аудиторські фірми; відділи податкових інспекцій, які займаються прийманням і перевіркою поточної звітності підприємств; квиткові каси; магазини; перукарні; телефонні станції і т.п.

Основними компонентами системи масового обслуговування будь-якого виду є вхідний потік вимог або заявок на обслуговування; дисципліна черги; механізм обслуговування.

Вхідний потік вимог. Для опису вхідного потоку потрібно задати імовірнісний закон, що визначає послідовність моментів надходження вимог на обслуговування і вказати кількість таких вимог в кожному черговому надходженні. При цьому, як правило, оперують поняттям «імовірнісний розподіл моментів надходження вимог». Тут можуть надходити як одиничні, так і групові вимоги (надходять групами). В останньому випадку звичайно йдеться про систему обслуговування з паралельно-груповим обслуговуванням.

Дисципліна черги - це важливий компонент системи масового обслуговування, він визначає принцип, згідно з яким заявки, що надходять на вхід обслуговуючої системи, підключаються із черги до процедури обслуговування. Найчастіше використовують дисципліни черги, обумовлені наступними правилами:

- перший прийшов - перший обслуговується;
- прийшов останнім - обслуговується першим;
- випадковий відбір заявок;
- відбір заявок за критерієм пріоритетності;
- обмеження часу очікування моменту настання обслуговування (має місце черга з обмеженим часом очікування обслуговування, що асоціюється з поняттям «припустима довжина черги»).

Механізм обслуговування визначається характеристиками самої процедури обслуговування і структурою обслуговуючої системи. До характеристик процедури обслуговування відносяться: тривалість обслуговування і кількість вимог, що задовольняються в результаті виконання кожної такої процедури. Для аналітичного опису характеристик обслуговування оперують поняттям «імовірнісний розподіл часу обслуговування». Слід зазначити, що час обслуговування заявки залежить від характеру самої заявки або вимог клієнта і від стану та можливостей обслуговуючої системи. Структура СМО визначається кількістю і взаємним розташуванням каналів обслуговування (механізмів, приладів і т.п.). Система може мати не один канал обслуговування, а декілька. Така система здатна обслуговувати одночасно декілька вимог. У цьому разі всі канали обслуговування пропонують ті самі послуги і, отже, можна стверджувати, що має місце паралельне обслуговування. Система може складатися з декількох різнотипних каналів обслуговування, через які повинна пройти кожна обслуговувана заявка, тобто в обслуговуючій системі процедури обслуговування заявок реалізуються послідовно. Механізм обслуговування визначає характеристики вихідного (обслуженого) потоку заявок.

Предметом теорії масового обслуговування є побудова математичних моделей, що зв'язують задані умови роботи СМО (число каналів, їхню продуктивність, характер потоку заявок і т.п.) з показниками ефективності СМО, що описують її здатність справлятися з потоком заявок. Показниками ефективності СМО є: середнє число заявок, що обслуговуються в одиницю часу; середнє число заявок у черзі; середній час очікування обслуговування; середнє число зайнятих каналів; імовірність відмови в обслуговуванні без очікування; імовірність того, що число заявок в черзі перевищить певне значення та ін.

СМО ділять на два основних класи: СМО з відмовами і СМО з очікуванням (з чергою). У СМО з відмовами заявка, що надійшла в момент, коли всі канали зайняті, отримує відмову, залишає СМО і надалі не бере участі в процесі обслуговування (наприклад, заявка на телефонну розмову в момент, коли всі канали зайняті, отримує відмову і залишає СМО не обслуженою). У СМО з очікуванням заявка, що прийшла в момент, коли всі канали зайняті, не йде, а стає в чергу на обслуговування.

СМО з очікуванням підрозділяються на різні види залежно від того, як організована черга: з обмеженою або необмеженою довжиною черги, з обмеженим часом очікування і т.п.

Математичний аналіз СМО істотно полегшується, якщо процес її роботи - марківський.

Поняття марківського випадкового процесу

Випадковий характер потоку заявок, а також, у загальному випадку, і тривалості обслуговування $t_{об}$ приводить до того, що в системі масового обслуговування протікає випадковий процес.

Випадковий процес, що відбувається в системі, називається марківським, якщо для кожного моменту часу t_0 імовірність будь-якого стану системи в майбутньому залежить тільки від її стану в даний момент часу t_0 і не залежить від того, яким способом система прийшла в цей стан.

На практиці марківські процеси в чистому вигляді не зустрічаються, але нерідко доводиться мати справу з процесами, для яких впливом передісторії можна зневажити, і для їх вивчення можна застосовувати марківські моделі.

Велике значення мають марківські випадкові процеси з дискретними станами і безперервним часом. Процес називається процесом з **дискретними станами**, якщо його можливі стани S_1, S_2, ...S_n можна заздалегідь перелічити, а перехід системи зі стану в стан відбувається миттєво (стрибком).

Процес називається процесом з **безперервним часом**, якщо моменти можливих переходів системи зі стану в стан не фіксовані заздалегідь, а випадкові.

Процес роботи СМО являє собою випадковий процес з дискретними станами і безперервним часом. Це означає, що стан СМО міняється стрибком у випадкові моменти появи якихось подій (наприклад, приходу нової заявки, закінчення обслуговування і т.п.).

У випадку немарківських процесів задачі дослідження систем масового обслуговування значно ускладнюються і вимагають застосування статистичного моделювання і чисельних методів з використанням ЕОМ.

При аналізі випадкових процесів з дискретними станами зручно користуватися спеціальною геометричною схемою - графом станів. Звичайно стани системи зображуються прямокутниками (або кружками), а можливі переходи з одного стану в інший - стрілками (орієнтованими дугами), що з'єднують стани.

Для математичного опису марківського випадкового процесу з дискретними станами і безперервним часом, що протікає в СМО, познайомимся з одним з важливих понять - поняттям потоку подій.

Потоки подій

Під потоком подій розуміється послідовність однорідних подій, що йдуть одна за іншою в якісь випадкові моменти часу (наприклад, потік викликів на телефонній станції, потік відмов ЕОМ, потік покупців і т.п.). Потік характеризується інтенсивністю λ - частотою появи подій або середнім числом подій, що надходять у СМО в одиницю часу.

Потік подій називається **регулярним**, якщо події йдуть одна за іншою через рівні проміжки часу.

Потік подій називається **стаціонарним**, якщо його імовірнісні характеристики не залежать від часу. Зокрема, інтенсивність стаціонарного потоку є величина постійна: $\lambda(t)=\lambda$.

Потік подій називається **потоком без післядії**, якщо для будь-яких двох непересічних ділянок часу τ_1 і τ_2 - число подій, що потрапляють на одну з них, не залежить від числа подій, що потрапляють на інші.

Потік подій називається **ординарним**, якщо імовірність попадання на малу (елементарну) ділянку часу Δt двох і більше подій зневажливо мала в порівнянні з імовірністю попадання однієї події.

Нагадаємо, що потік подій називається найпростішим, якщо він одночасно стаціонарний, ординарний і не має післядії. Назва "найпростіший" пояснюється тим, що СМО з найпростішими потоками має найбільш простий математичний опис. Для найпростішого потоку число подій m, що потрапляють на довільну ділянку часу τ, розподілено за законом Пуассона

$$P_m(\tau)=\frac{(\lambda\tau)^m}{m!}e^{-\lambda\tau},$$

для якого математичне сподівання випадкової величини m дорівнює її дисперсії:

$$a=\sigma^2=\lambda\tau.$$

Зокрема, імовірність того, що за час τ не відбудеться жодної події (m=0), дорівнює

$$P_0=e^{-\lambda\tau}.$$

Інтервал часу між довільними двома сусідніми подіями найпростішого потоку T розподілений за експонентним законом:

$$F(t)=P(T<t)=1-e^{-\lambda t}.$$

Щільність імовірності випадкової величини T є похідна її функції розподілу, тобто

$$f(t)=F'(t)=\lambda e^{-\lambda t}.$$

Математичне сподівання є величиною, зворотною λ, і дорівнює середньому квадратичному відхиленню випадкової величини T:

$$m=\sigma=\frac{1}{\lambda}.$$

Найважливіша властивість експонентного розподілу (притаманна тільки експонентному розподілу) полягає в наступному:

якщо проміжок часу, розподілений за експонентним законом, уже тривав якийсь час τ, то це ніяк не впливає на закон розподілу частини, що залишилася, проміжку (T-τ): він буде таким же, як і закон розподілу всього проміжку T.

Інакше кажучи, для інтервалу часу T між двома послідовними сусідніми подіями потоку, що має експонентний розподіл, будь-які відомості про те, скільки часу протікав цей інтервал, не впливають на закон розподілу частини, що залишилася. Ця властивість експонентного закону являє собою, по суті, інше формулювання для "відсутності післядії" - основної властивості найпростішого потоку.

В якості «міри випадковості» невід’ємної випадкової величини використовують коефіцієнт варіації:

$$V = \frac{\sigma}{m_t}.$$

Для **експонентного** розподілу, а отже для найпростішого потоку, коефіцієнт варіації дорівнює одиниці. Для регулярного потоку, в якого інтервал між подіями невипадковий і σ=0, коефіцієнт варіації дорівнює нулю. Для більшості реальних потоків подій коефіцієнт варіації інтервалів між подіями лежить у межах

$$0 \le V \le 1.$$

Він може служити мірою ступеня регулярності потоку. Чим ближче V до нуля, тим регулярніше потік. Найпростіший потік найменш регулярний із всіх потоків подій.

Рівняння для імовірностей станів

Як приклад розглянемо одноканальну СМО. Очевидно, що така СМО може перебувати в одному з двох станів:

S_0 – канал вільний;
S_1 – канал зайнятий.

Якщо всі потоки подій, що переводять систему з одного стану в інший, найпростіші, то процес, що протікає в системі, буде марківським. У цьому випадку його можна описати звичайними диференціальними рівняннями.

Побудуємо граф станів для одноканальної СМО (рис. 11.8).

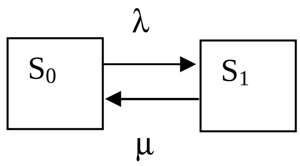


Рис. 11.8 - Спрямований граф одноканальної СМО з відмовами

Зі стану S_0 у стан S_1 систему переводить потік заявок, інтенсивність якого λ, а зі стану S_1 у стан S_0 систему переводить потік обслуговувань, інтенсивність якого μ. Нехай імовірності станів системи будуть p_0 і p_1 відповідно. Очевидно, що для будь-якого моменту часу сума імовірностей станів є імовірністю достовірної події

$$p_0(t) + p_1(t) = 1. \qquad (11.22)$$

Якщо є граф станів, можна визначити імовірності станів системи $p_0(t)$ і $p_1(t)$ як функції часу. Для цього складають і вирішують рівняння Колмогорова - диференціальні рівняння, в яких невідомими функціями є імовірності станів.

Оскільки потоки заявок і обслуговувань – найпростіші, інтенсивність λ =

const, і час між заявками розподілений за експонентним законом $f(t)= \lambda * e^{-\lambda t}$. Для процесу обслуговувань також μ= const, тривалість обслуговування $t_{об}$ розподілена за експонентним законом, тому $f(t_{об})= \mu * e^{-\mu t}$.

Розглянемо одну з імовірностей станів, наприклад, $p_0(t)$. Це імовірність того, що в момент часу t система перебуватиме в стані S_0. Знайдемо імовірність $p_0(t+\Delta t)$, тобто імовірність того, що в момент часу $(t+\Delta t)$ канал перебуває у вільному стані. У момент часу $(t+\Delta t)$ система перебуватиме в стані S_0 в двох випадках:

а) якщо в момент часу t система перебувала у стані S_0 і за час Δt вона не вийшла із цього стану;

б) якщо система в момент часу t була в стані S_1 і за час Δt перейшла в стан S_0.

Знайдемо імовірність варіанта а). Імовірність, що система в момент часу t перебувала у стані S_0 дорівнює $p_0(t)$. Імовірність того, що за час Δt не надійшло ні однієї заявки визначиться за законом Пуассона

$$\frac{(\lambda \Delta t)^0}{0!} e^{-\lambda \Delta t} = e^{-\lambda \Delta t} \qquad (11.23)$$

розкладання в ряд Тейлора дає приблизно

$$e^{-\lambda \Delta t} = 1-\lambda \Delta t+\lambda^2 \Delta t^{2/2} - \ldots \approx 1-\lambda \Delta t.$$

Таким чином, імовірність дотримання умови "а)" за теоремою множення дорівнює:

$$p_0(t) * (1-\lambda \Delta t)\,. \qquad (11.24)$$

Знайдемо імовірність варіанту б). Імовірність, що система в момент часу t перебувала у стані S_1 дорівнює $p_1(t)$. Імовірність того, що за час Δt канал звільнився, визначиться за законом Пуассона через імовірність протилежної події

$$1-\frac{(\mu \Delta t)^0}{0!} e^{-\mu \Delta t} = 1 - e^{-\mu \Delta t} \approx \mu \Delta t \qquad (11.25)$$

Тоді імовірність дотримання умови "б)" за теоремою множення

$$p_1(t) * \mu \Delta t. \qquad (11.26)$$

Таким чином, імовірність того, що в момент часу $(t+\Delta t)$ система перебуватиме в стані S_0 (канал вільний), визначиться за теоремою додавання імовірностей

$$p_0(t+\Delta t) = p_0(t) * (1-\lambda \Delta t) + p_1(t) * \mu \Delta t. \qquad (11.27)$$

Перенесемо $p_0(t)$ у ліву частину, розділимо на Δt і в границі при $\Delta t \to 0$ отримаємо

$$\frac{dp_0(t)}{dt} = -p_0(t)\lambda + p_1(t)\mu \qquad (11.28)$$

Диференціальне рівняння для стану S_1 дістанемо, міркуючи аналогічно:

$$\frac{dp_1(t)}{dt} = p_0(t)\lambda - p_1(t)\mu\,. \qquad (11.29)$$

Спрямований граф, наведений на рис. 11.8, ілюструє процес зміни станів. Його вершини відповідають станам, а ребра - можливим переходам з одного стану в інший.

Якщо є спрямований граф станів, то систему диференціальних рівнянь для імовірностей станів p_k (k=0,1,2,…) можна записати, користуючись наступним правилом:

в лівій частині кожного диференціального рівняння знаходиться похідна імовірності k-го стану, а в правій - стільки доданків, скільки ребер зв'язано безпосередньо з даним станом;

якщо ребро закінчується в даному стані, то доданок має знак плюс, якщо починається з даного стану, - мінус;

кожен доданок дорівнює добутку інтенсивності потоку подій, що переводить елемент або систему за даному ребром в інший стан, на імовірність того стану, з якого починається ребро.

Щоб вирішити рівняння Колмогорова і знайти імовірності станів, необхідно насамперед задатися початковими умовами. Наприклад, у момент часу t = 0 система перебувала в стані S_1. Тоді початкові умови $p_0(0) = 1$, $p_1(0) = 0$. Подібні рівняння можна вирішувати аналітично або чисельно - вручну або на ЕОМ. У результаті вирішення визначаються імовірності станів як функції часу:

$$p_0(t) = \frac{\mu}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t},$$

$$p_1(t) = \frac{\lambda}{\lambda+\mu} + \frac{\lambda}{\lambda+\mu} e^{-(\lambda+\mu)t}. \qquad (11.30)$$

При $t \to \infty$ в системі встановлюється граничний стаціонарний режим, в ході якого система випадковим чином міняє свої стани, але їхні імовірності вже не залежать від часу. Ці імовірності називаються **фінальними імовірностями станів**. Фінальну імовірність стану розуміють як середній відносний час перебування системи в цьому стані. Наприклад, якщо система має два стани, фінальні імовірності яких 0,3 і 0,7, то це означає, що система в середньому 30% часу проводить у стані S_0.і 70% - у стані S_1. Можна оцінити середню ефективність роботи системи. Нехай в стані S_0 вона дає нульовий дохід, а в стані S_1 дохід становить 15 грн. Тоді

$$E = 0{,}3*0+0{,}7*15=10{,}5 \text{ грн.}$$

Для обчислення фінальних імовірностей в системі диференціальних рівнянь необхідно покласти похідні імовірностей рівними нулю і розв'язати отриману систему алгебраїчних рівнянь. Можна також із графа станів відразу записати систему алгебраїчних рівнянь для фінальних імовірностей станів. Для цього користуються правилом:

- ліворуч стоїть фінальна імовірність стану p_k, помножена на сумарну інтенсивність всіх потоків, що переводять систему з даного стану в інші;

- праворуч стоїть сума добутків інтенсивностей всіх потоків, що переводять систему в даний стан на імовірності тих станів, з яких ці потоки виходять.

Визначимо характеристики одноканальної системи масового обслуговування з відмовами. Для такої системи імовірність p_0 є відносною пропускною здатністю системи q. Дійсно, p_0 - імовірність того, що в момент t канал вільний і заявка, що прийшла, буде обслуженою. З формули (11.30) при $t \to \infty$ дістанемо:

$$q = p_0 = \frac{\mu}{\lambda+\mu} \quad . \tag{11.31}$$

Знаючи відносну пропускну здатність q, легко знайти абсолютну. Абсолютна пропускна здатність A - це середнє число заявок, які може обслужити система масового обслуговування в одиницю часу:

$$A = \lambda q = \frac{\lambda\mu}{\lambda+\mu}. \tag{11.32}$$

Імовірність відмови в обслуговуванні заявки дорівнюватиме імовірності стану «канал зайнятий»:

$$P_{відм} = p_1 = 1 - p_0 = 1 - \frac{\mu}{\lambda+\mu} = \frac{\lambda}{\lambda+\mu}. \tag{11.33}$$

Дана величина $P_{відм}$ може бути інтерпретована як середня частка не обслужених заявок серед тих, що прийшли.

Приклад 11.4. Нехай одноканальна СМО з відмовами являє собою один пост щоденного обслуговування для мийки автомобілів. Заявка - автомобіль, що прибув в момент, коли пост зайнятий, - одержує відмову в обслуговуванні. Інтенсивність потоку автомобілів λ = 1,0 (автомобіль на годину). Середня тривалість обслуговування - 1,8 години. Потік автомобілів і потік обслуговувань є найпростішими. Потрібно визначити відносну пропускну здатність q; абсолютну пропускну здатність A; імовірність відмови $P_{відм}$, а також зрівняти фактичну пропускну здатність СМО з номінальною, котра була б, якби кожен автомобіль обслуговувався точно 1,8 години і автомобілі надходили один за одним без перерви.

Розв'язання: Визначимо інтенсивність потоку обслуговувань:

$$\mu = \frac{1}{t_{об}} = \frac{1}{1{,}8} = 0{,}555.$$

Обчислимо відносну пропускну здатність:

$$q = \frac{\mu}{\mu+\lambda} = \frac{0{,}555}{1+0{,}555} = 0{,}356 \quad .$$

Величина q=0,356 означає, що система в середньому буде обслуговувати 35,6% автомобілів, які прибувають на пост.

Абсолютну пропускну здатність знайдемо за формулою

$$A = q\lambda = 0{,}356 * 1 = 0{,}356 \quad .$$

Це означає, що система здатна обслужити в середньому 0,356 автомобілів на годину.

Імовірність відмови:

$$P_{відм} = 1 - q = 1 - 0{,}356 = 0{,}644 \quad .$$

Це означає, що 64,4% прибулих автомобілів одержать відмову в обслуговуванні.

Визначимо номінальну пропускну здатність системи:

$$A = \frac{1}{t_{об}} = \frac{1}{1{,}8} = 0{,}555 \text{ (автомобілів на годину).}$$

Виявляється, що $A_{ном}$ в 1,5 раза більше, ніж фактична пропускна здатність, обчислена з урахуванням випадкового характеру потоку заявок і часу обслуговування ($\frac{A_{ном}}{A} = \frac{0,555}{0,356} = 1,56$).

Багатоканальна СМО з відмовами

Розглянемо функціонування системи з відмовами (класична задача Ерланга). Нехай є n-канальна система, яку можна представити як фізичну систему з кінцевою множиною станів: S_0 – всі канали вільні, S_1 – зайнятий рівно один канал, S_k – зайнято рівно k каналів,…, S_n–зайняті всі n каналів. Граф станів системи показаний на рис. 11.9:

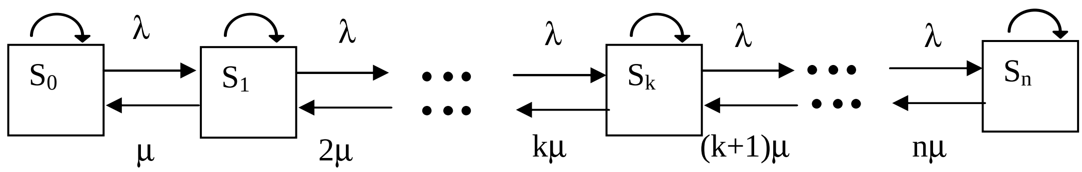


Рис. 11.9 - Граф станів СМО з відмовами

У верхніх стрілок графа інтенсивності λ, тому що той самий потік заявок переводить систему з одного стану в інший. У нижніх стрілок – інтенсивність обслуговувань μ. У стані S_1 зайнятий один канал, і в стан S_0 система перейде, коли він звільниться. У стані S_2 зайняті два канали, і в стан S_1 система перейде, коли звільняться перший або другий канали, інтенсивність переходу дорівнює $\mu+\mu=2\mu$, і т.д.

Імовірності станів системи $p_k(t)$ для будь-якого моменту часу t можна визначити в такий спосіб. Нехай потоки заявок на обслуговування і звільнень є найпростішими, інтервал між заявками розподілений за експонентним законом з параметром λ.

Складемо рівняння для $p_k(t)$ ($1 \le k \le n$). Зафіксуємо момент часу t і знайдемо імовірність $p_k(t+\Delta t)$ того, що в момент часу $t+\Delta t$ система буде знаходитись в стані S_k. Ця імовірність обчислюється як імовірність суми трьох подій A, B і C (по числу стрілок, спрямованих в стан S_k на рис. 11.9): подія A={в момент t система перебувала в стані S_k і за час Δt заявка не надійшла і жоден канал не звільнився}; подія B={в момент t система перебувала у стані S_{k-1}, і за час Δt надійшла одна заявка}; подія C={в момент t система перебувала у стані S_{k+1}, і за час Δt канал звільнився}. За теоремою додавання імовірностей маємо:

$$p_k(t+\Delta t) = P(A) + P(B) + P(C) \qquad (11.34)$$

Імовірність події A, тобто того, що за час Δt не надійшла жодна заявка і жоден канал не звільнився, знайдемо відповідно до теореми множення імовірностей:

$$e^{-\lambda\Delta t}\left(e^{-\mu\Delta t}\right)^k = e^{-(\lambda+\mu k)\Delta t}$$

або, зневажаючи величинами малих порядків, маємо:

$$e^{-(\lambda+\mu k)\Delta t} \approx 1-(\lambda+\mu k)\Delta t,$$

тоді

$$P(A) = p_k(t)[1-(\lambda+\mu k)\Delta t] \tag{11.35}$$

Аналогічно визначимо імовірності подій В і С:

$$P(B) = p_{k-1}(t)\lambda\Delta t \tag{11.36}$$

$$P(C) = p_{k+1}(t)(k+1)\mu\Delta t. \tag{11.37}$$

Підставивши (11.35),(11.36) і (11.37) в (11.34), дістанемо

$$p_k(t+\Delta t) = p_k(t)[1-(\lambda+\mu k)\Delta t] + p_{k-1}(t)\lambda\Delta t + p_{k+1}(t)(k+1)\mu\Delta t.$$

Перенесемо $p_k(t)$ в ліву частину рівняння, розділивши на Δt і перейшовши до границі при $\Delta t \to 0$, отримаємо диференціальне рівняння для $p_k(t)$:

$$\frac{dp_k(t)}{dt} = \lambda p_{k-1}(t) - (\lambda + k\mu)p_k(t) + (k+1)\mu p_{k+1}(t)\ . \tag{11.38}$$

Аналогічно можна отримати диференціальні рівняння для імовірностей станів системи при k=0 і k=n:

$$\begin{cases} \dfrac{dp_0(t)}{dt} = -\lambda p_0(t) + \mu p_1(t) \\ \\ \dfrac{dp_k(t)}{dt} = \lambda p_{k-1}(t) - (\lambda + k\mu)p_k(t) + (k+1)\mu p_{k+1}(t) \\ \\ \dfrac{dp_n(t)}{dt} = \lambda p_{n-1}(t) - n\mu p_n(t). \end{cases} \tag{11.39}$$

Інтегрування системи рівнянь (11.39) при початкових умовах $p_0(0)$ =1, $p_1(0)$ =…=$p_n(0)$ =0 (в початковий момент часу всі канали вільні) дозволяє знайти імовірність кожного з $k=\overline{0,n}$ станів системи.

Імовірність $p_n(t)$ – це імовірність того, що заявка, яка прийшла в момент часу t, застане всі канали зайнятими, тобто одержить відмову. Імовірність $q(t)=1-p_n(t)$ є відносною пропускною здатністю системи. Для даного t це є відношення середнього числа обслужених за одиницю часу заявок до середнього числа поданих.

Замінивши в рівняннях Колмогорова (11.39) похідні імовірностей станів нульовими значеннями, отримаємо систему лінійних алгебраїчних рівнянь, що описують стаціонарний режим:

$$\begin{cases} -\lambda p_0 + \mu p_1 = 0 \\ \ldots \\ \lambda p_{k-1} - (\lambda + k\mu)p_k + (k+1)\mu p_{k+1} = 0 \\ \ldots \\ \lambda p_{n-1} - n\mu p_n = 0 \end{cases} \tag{11.40}$$

Вирішивши систему (11.40) відносно імовірностей, дістанемо для будь-якого k:

$$p_k = \frac{\lambda^k}{k!\mu^k} p_0 \qquad (11.41)$$

Позначимо відношення $\lambda/\mu=\alpha$ і назвемо його приведеною інтенсивністю потоку заявок, яка є числом заявок, що приходять за середній час обслуговування однієї заявки. При цьому формула (11.41) набуде вигляду

$$p_k = \frac{\alpha^k}{k!} p_0 . \qquad (11.42)$$

Звідки, врахувавши, що $\sum_{k=0}^{n} p_k = 1$, дістанемо:

$$\sum_{k=0}^{n} p_k = p_0 \sum_{k=0}^{n} \frac{\alpha^k}{k!} = 1 . \qquad (11.43)$$

Виразимо звідси p_0 і підставимо в (11.42)

$$p_k = \frac{\frac{\alpha^k}{k!}}{\sum_{k=0}^{n} \frac{\alpha^k}{k!}} , \qquad (11.44)$$

де $k = \overline{0,n}$.

Формула (11.44) називається формулою Ерланга. Вона дає граничний закон розподілу числа зайнятих каналів залежно від характеристик потоку заявок і продуктивності системи обслуговування.

Зазначимо, що незважаючи на те, що формули Ерланга в точності справедливі тільки для найпростішого потоку заявок, ними з відомим наближенням можна користуватися і для потоків з обмеженою післядією, а також для систем з очікуванням, коли строк очікування заявки в черзі малий в порівнянні з середнім часом обслуговування.

Визначимо характеристики функціонування багатоканальної СМО з відмовами у стаціонарному режимі:

- імовірність відмови:

$$P_{відм} = p_n = \frac{(\alpha)^n}{n!} * p_0 , \qquad (11.45)$$

тому що заявка одержує відмову, якщо приходить в момент, коли всі n каналів зайняті. Величина $P_{отк}$ характеризує повноту обслуговування вхідного потоку;

імовірність того, що заявка буде прийнята до обслуговування (вона ж - відносна пропускна здатність системи q, доповнює $P_{відм}$ до одиниці):

$$q = 1 - P_{відм} = 1 - \frac{(\alpha)^n}{n!} * p_0 , \qquad (11.46)$$

абсолютна пропускна здатність

$$A = \lambda q = \lambda(1 - P_{відм}) , \qquad (11.47)$$

середнє число каналів, зайнятих обслуговуванням ($\overline{k}$), наступне:

$$\overline{k} = \sum_{k=1}^{n} k p_k = \alpha(1 - P_{відм}) . \qquad (11.48)$$

Величина $\overline{k}$ характеризує ступінь завантаження СМО.

Приклад 11.5. Нехай n-канальна СМО являє собою обчислювальний центр з трьома (n = 3) взаємозамінними ПЕОМ для вирішення задач. Потік задач, що надходять на ОЦ, має інтенсивність λ = 1 задача на годину. Середня тривалість обслуговування $\bar{t}_{об}$ = 1,8 години. Потік заявок на вирішення задач і потік обслуговування цих заявок є найпростішими. Потрібно обчислити фінальні імовірності станів ОЦ; імовірності відмови в обслуговуванні заявки; відносну пропускну здатність ОЦ; абсолютну пропускну здатність ОЦ; середнє число зайнятих ПЕОМ, а також визначити, скільки додатково треба придбати ПЕОМ, щоб скоротити імовірність відмови в обслуговуванні заявки в 2 рази.

Розв'язання: Визначимо параметр потоку обслуговувань μ:

$$\mu = \frac{1}{\bar{t}_{об}} = \frac{1}{1,8} = 0,555\,.$$

Приведена інтенсивність потоку заявок $\alpha = \dfrac{1}{0,555} = 1,8\,.$

Фінальні імовірності станів знайдемо за формулами Ерланга (11.42)-(11.44):

$$p_1 = \frac{\alpha}{1!} * p_0 = \frac{1,8}{1!} = 1,8 * p_0\,,$$

$$p_2 = \frac{(\alpha)^2}{2!} * p_0 = 1,62 * p_0\,,$$

$$p_3 = \frac{(\alpha)^3}{3!} * p_0 = 0,97 * p_0\,,$$

$$p_0 = \frac{1}{\sum_{k=0}^{3} \frac{(\alpha)^k}{k!}} = \frac{1}{1 + 1,8 + 1,62 + 0,97} \approx 0,186\,,$$

$$p_1 = 1,8 * 0,186 = 0,344\,,$$

$$p_2 = 1,62 * 0,186 = 0,301\,,$$

$$p_3 = 0,97 * 0,186 = 0,18\,.$$

Імовірність відмови в обслуговуванні заявки

$$P_{відм} = p_3 = 0,18\,.$$

Відносна пропускна здатність ОЦ

$$q = 1 - P_{відм} = 1 - 0,18 = 0,82\,.$$

Тобто 18% заявок залишають систему не обслуженими.

Абсолютна пропускна здатність ОЦ

$$A = \lambda q = 1 * 0,82 = 0,82\,.$$

Середнє число зайнятих каналів - ПЕОМ

$$\bar{k} = \alpha(1 - P_{відм}) = 1,8 * (1 - 0,18) = 1,476\,.$$

Таким чином, при сталому режимі роботи СМО в середньому буде зайнято 1,5 комп'ютера з трьох - інші півтора будуть простоювати. Очевидно, що пропускну здатність ОЦ при даних λ і μ можна збільшити тільки за рахунок збільшення числа ПЕОМ.

Визначимо, скільки потрібно використати ПЕОМ, щоб скоротити імовірність відмови в обслуговуванні заявки в 2 рази, тобто щоб імовірність відмови у вирішенні задач не перевершувала 0,09. Для цього скористаємось формулою (11.44)

$$P_{відм} = \frac{(\alpha)^n}{n!} * p_0 .$$

Результати розрахунків зведемо в таблицю:

n	1	2	3	4	5
p_0	0,357	0,266	0,186	0,172	0,167
$P_{відм}$	0,643	0,367	0,18	0,075	0,026

Таким чином, щоб зменшити імовірність відмови в обслуговуванні заявки в 2 рази необхідно додати ще один комп'ютер.

Одноканальна СМО з очікуванням і обмеженою чергою

Якщо заявка, яка застала всі канали зайнятими, стає в чергу і чекає, поки звільниться який-небудь канал, причому час очікування не обмежений, то така система називається чистою системою з очікуванням. Якщо час очікування обмежений, то система називається системою змішаного типу. Обмеження можуть накладатися на час очікування заявки в черзі, на число заявок в черзі. Змішані системи деяких типів можуть бути описані диференціальними рівняннями аналогічними рівнянням Ерланга. При цьому час очікування заявки в черзі, як і час її обслуговування, вважається випадковим, розподіленим за експонентним законом. Це дозволяє вважати процеси, що протікають в системі, марківськими.

Система масового обслуговування має один канал. Вхідний потік заявок на обслуговування - найпростіший потік з інтенсивністю λ. Інтенсивність потоку обслуговування дорівнює μ. Тривалість обслуговування - випадкова величина, підлегла експонентному закону розподілу. Заявка, що надійшла в момент, коли канал зайнятий, стає в чергу і очікує обслуговування.

Припустимо, що незалежно від того, скільки вимог надходить на вхід обслуговуючої системи, дана система (черга плюс клієнти, які обслуговуються) не може вмістити більше N вимог (заявок), тобто клієнти, що не потрапили в очікування, змушені обслуговуватися в іншому місці.

Граф станів СМО в цьому разі має вигляд, показаний на рис. 11.10.

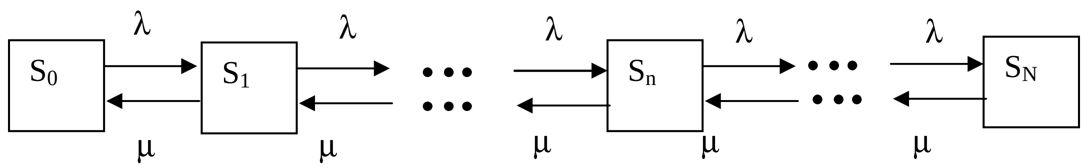


Рис. 11.10 - Граф станів одноканальної СМО з очікуванням

Стани СМО мають наступну інтерпретацію:

S_0 - «канал вільний»;

S_1 - «канал зайнятий» (черги немає);

S_2 - «канал зайнятий» (одна заявка стоїть в черзі);

…………………………………………………......

S_n - «канал зайнятий» (n -1 заявок стоять в черзі);

S_N - «канал зайнятий» (N - 1 заявок стоять в черзі).

Стаціонарний процес в даній системі буде описуватися наступною системою алгебраїчних рівнянь:

$$\begin{cases} -\alpha p_0 + p_1 = 0, ...n = 0 \\ -(1-\alpha)p_n + p_{n+1} + \alpha p_{n-1} = 0, ...0 < n < N \\ -p_N + \alpha P_{N-1} = 0, ...n = N \end{cases}, \tag{11.49}$$

де n - номер стану.

Рішення наведеної вище системи рівнянь для нашої моделі СМО має вигляд

$$p_n = \begin{cases} \left(\dfrac{1-\alpha}{1-\alpha^{N+1}}\right)\alpha^n, ...\alpha \neq 1, ...n = \overline{0, N} \\ \dfrac{1}{N+1}, ...\alpha = 1. \end{cases} \tag{11.50}$$

$$p_0 = \frac{1-\alpha}{1-\alpha^{N+1}} \ . \tag{11.51}$$

Тоді

$$p_n = \begin{cases} p_0\alpha^n, ...\alpha \neq 1, ...n = \overline{0, N} \\ \dfrac{1}{N+1}, ...\alpha = 1. \end{cases}.$$

Слід зазначити, що виконання умови стаціонарності $\alpha < 1$ для даної СМО не обов'язкове, оскільки число заявок, що допускаються в обслуговуючу систему, контролюється шляхом введення обмеження на довжину черги (яка не може перевищувати N-1), а не співвідношенням між інтенсивностями вхідного потоку, тобто не відношенням $\frac{\lambda}{\mu} = \alpha$.

Визначимо характеристики одноканальної СМО з очікуванням і обмеженою довжиною черги, рівною (N - 1):

- імовірність відмови в обслуговуванні заявки:

$$P_N = \begin{cases} \left(\dfrac{1-\alpha}{1-\alpha^{N+1}}\right)\alpha^n, ...\alpha \neq 1, \\ \dfrac{1}{N+1}, ...\alpha = 1. \end{cases} \tag{11.52}$$

- відносна пропускна здатність системи:

$$q = 1 - P_{відм} = \begin{cases} 1 - \left(\dfrac{1-\alpha}{1-\alpha^{N+1}}\right)\alpha^{n}, ...\alpha \neq 1, \\ 1 - \dfrac{1}{N+1}, ...\alpha = 1. \end{cases} \tag{11.53}$$

- абсолютна пропускна здатність:

$$A = \lambda q \tag{11.54}$$

- середнє число заявок, що перебувають в системі:

$$L_s = \sum_{n=0}^{N} n * p_n = \begin{cases} \dfrac{\alpha\left[1-(N+1)\alpha^{N} + N\alpha^{N+1}\right]}{(1-\alpha)\left(1-\alpha^{N+1}\right)}, ...\alpha \neq 1, \\ \dfrac{N}{2}, ...\alpha = 1. \end{cases} \tag{11.55}$$

- середній час перебування заявки в системі:

$$W_s = \frac{L_s}{\lambda(1-p_N)} \tag{11.56}$$

- середня тривалість перебування клієнта (заявки) в черзі:

$$W_q = W_s - \frac{1}{\mu} \tag{11.57}$$

- середнє число заявок (клієнтів) в черзі (довжина черги):

$$L_q = \lambda(1-p)W_q \tag{11.58}$$

Приклад 11.6. Спеціалізований пост діагностики являє собою одноканальну СМО. Число стоянок для автомобілів, що очікують проведення діагностики, обмежене і дорівнює 3 ((N - 1) = 3). Якщо всі стоянки зайняті, тобто в черзі вже перебувають три автомобілі, то черговий автомобіль, який прибув на діагностику, у чергу на обслуговування не стає. Потік автомобілів, які прибувають на діагностику, має інтенсивність λ = 0,85 (автомобіля на годину). Час діагностики автомобіля в середньому дорівнює 1,05 години. Потрібно визначити характеристики поста діагностики, що працює у стаціонарному режимі.

Розв'язання: Параметр потоку обслуговувань автомобілів:

$$\mu = \frac{1}{\overline{\overline{t_{об}}}} = \frac{1}{1{,}05} = 0{,}952.$$

Приведена інтенсивність потоку автомобілів визначається як відношення інтенсивностей λ і μ, тобто $\alpha = \dfrac{\lambda}{\mu} = \dfrac{0{,}85}{0{,}952} = 0{,}893$.

Обчислимо фінальні імовірності станів системи:

$$p_0 = \frac{1-\alpha}{1-\alpha^{N+1}} = \frac{1-0{,}893}{1-0{,}893^5} \approx 0{,}248,$$

$$p_1 = \alpha p_0 = 0{,}893 * 0{,}248 \approx 0{,}221,$$

$$p_2 = \alpha^2 p_0 = 0{,}893^2 * 0{,}248 \approx 0{,}198,$$

$$p_3 = \alpha^3 p_0 = 0{,}893^3 * 0{,}248 \approx 0{,}177,$$

$$p_4 = \alpha^4 p_0 = 0{,}893^4 * 0{,}248 \approx 0{,}158.$$

Імовірність відмови в обслуговуванні автомобіля:

$$P_{відм} = p_4 = \alpha^4 p_0 \approx 0{,}158 \ .$$

Відносна пропускна здатність поста діагностики:

$$q = 1 - P_{відм} = 1 - 0{,}158 = 0{,}842.$$

Абсолютна пропускна здатність поста діагностики

$$A = q*\lambda = 0{,}842 * 0{,}85 = 0{,}716 \text{ (автомобіля на годину).}$$

Середнє число автомобілів, що перебувають на обслуговуванні і в черзі (тобто в системі масового обслуговування):

$$L_s = \frac{\alpha\left[1-(N+1)\alpha^N + N\alpha^{N+1}\right]}{(1-\alpha)\left(1-\alpha^{N+1}\right)} = \frac{0{,}893\left[1-(4+1)(0{,}893)^4 + 4*0{,}893^5\right]}{(1-0{,}893)\left(1-0{,}893^5\right)} = 1{,}77 \ .$$

Середній час перебування автомобіля в системі:

$$W_s = \frac{L_s}{\lambda(1-p_N)} = \frac{1{,}77}{0{,}85*(1-0{,}158)} = 2{,}473 \text{ години.}$$

Середня тривалість перебування заявки в черзі на обслуговування:

$$W_q = W_s - \frac{1}{\mu} = 2{,}473 - \frac{1}{0{,}952} = 1{,}423 \text{ години.}$$

Середнє число заявок в черзі (довжина черги):

$$L_q = \lambda(1-P_N)W_q = 0{,}85*(1-0{,}158)*1{,}423 = 1{,}02.$$

Роботу розглянутого поста діагностики можна вважати задовільною, тому що пост діагностики не обслуговує автомобілі в середньому в 15,8% випадків ($P_{відм} = 0{,}158$).

Одноканальна СМО з очікуванням і необмеженою чергою

Розглянемо одноканальну СМО з очікуванням без обмеження на місткість блоку очікування (тобто $N \to \infty$). Решта умов функціонування СМО залишаються без змін.

Стаціонарний режим функціонування такої СМО існує при $t \to \infty$ для будь-якого n = 0, 1, 2, ... і коли $\lambda < \mu$. Система алгебраїчних рівнянь, що описують роботу СМО при цьому має вигляд

$$\begin{cases} -\lambda p_0 + \mu p_1 = 0, \ldots\ldots. n = 0 \\ \lambda p_{n-1} + \mu p_{n+1} - (\lambda + \mu) p_n = 0, \ldots\ldots n > 0 \end{cases} . \qquad (11.59)$$

Рішення даної системи рівнянь має вигляд

$$p_n = (1-\alpha)\alpha^n, \ n = 0, 1, 2, \ldots \qquad (11.60)$$

де $\alpha < 1$.

Характеристики одноканальної СМО з очікуванням, без обмеження на довжину черги, такі:

- середнє число клієнтів, які перебувають в системі (заявок) на обслуговування:

$$L_s = \sum_{n=0}^{\infty} n p_n = \frac{\alpha}{1-\alpha} \qquad (11.61)$$

- середня тривалість перебування клієнта в системі:

$$W_s = \frac{L_s}{\lambda} = \frac{1}{\mu(1-\alpha)} \qquad (11.62)$$

- середнє число клієнтів у черзі на обслуговування:

$$L_q = L_s - \alpha = \frac{\alpha^2}{1-\alpha} \qquad (11.63)$$

- середня тривалість перебування клієнта в черзі:

$$W_q = \frac{L_q}{\lambda} = \frac{\alpha}{\mu(1-\alpha)} \qquad (11.64)$$

Приклад 11.7. Повернемося до прикладу 11.6, де мова йде про функціонування поста діагностики. Нехай розглянутий пост діагностики має в своєму розпорядженні необмежену кількість площадок для стоянки автомобілів, які прибувають на обслуговування, тобто довжина черги не обмежена. Треба визначити фінальні значення імовірнісних характеристик.

Розв'язання: Параметр потоку обслуговувань μ і приведена інтенсивність потоку автомобілів α визначені в прикладі 4.2:

$$\mu = 0{,}952; \ \alpha = 0{,}893.$$

Обчислимо граничні імовірності системи за формулами

$$p_0 = 1-\alpha = 1-0{,}893 = 0{,}107,$$

$$p_1 = (1-\alpha)\alpha = (1-0{,}893)*0{,}893 = 0{,}096,$$

$$p_2 = (1-\alpha)\alpha^2 = (1-0{,}893)*0{,}893^2 = 0{,}085,$$

$$p_3 = (1-\alpha)\alpha^3 = (1-0{,}893)*0{,}893^3 = 0{,}076,$$

$$p_4 = (1-\alpha)\alpha^4 = (1-0{,}893)*0{,}893^4 = 0{,}068,$$

$$p_5 = (1-\alpha)\alpha^5 = (1-0{,}893)*0{,}893^5 = 0{,}061$$

і т.д.

Слід зазначити, що $p_0(t)$ визначає частку часу, протягом якого пост діагностики вимушено не діє (простоює). У нашому прикладі вона становить 10,7%, тому що $p_0(t) = 0{,}107$.

Середнє число автомобілів, які перебувають в системі (на обслуговуванні і в черзі):

$$L_s = \frac{\alpha}{1-\alpha} = \frac{0{,}893}{1-0{,}893} = 8{,}346 \text{ од.}$$

Середня тривалість перебування клієнта в системі:

$$W_s = \frac{1}{\mu(1-\alpha)} = \frac{1}{0{,}952(1-0{,}893)} = 8{,}766 \text{ години.}$$

Середнє число автомобілів в черзі на обслуговування:

$$L_q = \frac{\alpha^2}{1-\alpha} = \frac{0{,}893^2}{1-0{,}893} = 7{,}483 \ .$$

Середня тривалість перебування автомобіля в черзі:

$$W_q = \frac{\alpha}{\mu(1-\alpha)} = \frac{0{,}893}{0{,}952*(1-0{,}893)} = 7{,}766 \text{ години.}$$

Відносна пропускна здатність системи:

$$q = 1,$$

тобто кожна заявка, що прийшла в систему, буде обслужена.

Абсолютна пропускна здатність:

$$A = \lambda q = 0,85 * 1 = 0,85.$$

Слід зазначити, що підприємство, яке здійснює діагностику автомобілів, насамперед, цікавить кількість клієнтів, що відвідаютьє пост діагностики при знятті обмеження на довжину черги.

Допустимо, в первісному варіанті кількість місць для стоянки автомобілів, які прибувають, дорівнює трьом (у прикладі 11.6). Частота m виникнення ситуацій, що коли прибуває на пост діагностики автомобіль не має можливості приєднатися до черги

$$m = \lambda^* p\ .$$

У нашому прикладі при $N = 3 + 1 = 4$ і $\alpha = 0,893$,

$$m = \lambda^* p\ \alpha^4 = 0,85*0,284*0,893^4 = 0,134 \text{ автомобіля на годину.}$$

При 12-годинному режимі роботи поста діагностики це еквівалентно тому, що пост діагностики в середньому за зміну (день) буде втрачати 12 * 0,134 = 1,6 автомобіля.

Зняття обмеження на довжину черги дозволяє збільшити кількість обслужених клієнтів у нашому прикладі в середньому на 1,6 автомобіля за зміну (12 годин роботи). Ясно, що рішення щодо розширення площі для стоянки автомобілів повинне ґрунтуватися на оцінці економічного збитку, який обумовлений втратою клієнтів при наявності всього трьох місць для стоянки цих автомобілів, і витрат, необхідних для розширення площі для стоянки.

Багатоканальна СМО з очікуванням і необмеженою чергою

Процес обслуговування в багатоканальній СМО з очікуванням характеризується наступним: вхідний і вихідний потоки є пуассонівськими з інтенсивностями λ і μ відповідно. Система має S каналів обслуговування. Середня тривалість обслуговування одного клієнта дорівнює $\frac{1}{\mu}$.

У сталому режимі функціонування багатоканальної СМО з очікуванням і необмеженою чергою можна описати за допомогою системи алгебраїчних рівнянь:

$$\begin{cases} 0 = \lambda p_{n-1} - (\lambda + n\mu) p_n + (n+1)\mu p_{n+1}, \ldots\ldots\ldots при \ldots\ldots 1 \le n < S \\ 0 = \lambda p_{n-1} - (\lambda + S\mu) p_n + S\mu p_{n+1}, \ldots\ldots\ldots при \ldots\ldots n \ge S. \end{cases} \tag{11.65}$$

Рішення системи рівнянь (11.65) має вигляд

$$\begin{cases} p_n = \dfrac{\alpha^n}{n!} p_0, \ldots\ldots\ldots при \ldots\ldots 1 \le n < S & (11.66) \\ p_n = \dfrac{\alpha^n}{S! S^{n-S}} p_0, \ldots\ldots\ldots при \ldots\ldots n \ge S. & (11.67) \end{cases}$$

де

$$p_0 = \left[\sum_{n=0}^{S-1} \frac{\alpha^n}{n!} + \frac{\alpha^S}{S!\left(1 - \frac{\alpha}{S}\right)} \right]^{-1} . \qquad (11.68)$$

Рішення буде дійсним, якщо виконується умова $\frac{\alpha}{S} < 1$.

Імовірнісні характеристики функціонування у стаціонарному режимі багатоканальної СМО з очікуванням і необмеженою чергою визначаються за наступними формулами:

- імовірність того, що в системі перебуває n клієнтів на обслуговуванні, визначається за формулами (11.66) і (11.67);

- середнє число клієнтів в черзі на обслуговування

$$L_q = \frac{S\alpha}{(S-\alpha)^2} * p_s \ ; \qquad (11.69)$$

- середнє число клієнтів, які перебувають в системі (заявок на обслуговування і в черзі)

$$L_s = L_q + \alpha ; \qquad (11.70)$$

- середня тривалість перебування клієнта (заявки на обслуговування) в черзі

$$W_q = \frac{L_q}{\lambda} ; \qquad (11.71)$$

- середня тривалість перебування клієнта в системі

$$W_s = W_q + \frac{1}{\mu} . \qquad (11.72)$$

Розглянемо приклади багатоканальної системи масового обслуговування з очікуванням.

Приклад 11.8. Механічна майстерня заводу з трьома дільницями (каналами) виконує ремонт об'єктів малої механізації. Потік несправних механізмів, які прибувають в майстерню, має інтенсивність λ = 2,5 механізми на добу, середній час ремонту одного механізма розподілено за експонентним законом і дорівнює $\bar{t}_{об}$ = 0,5 діб. Припустимо, що іншої майстерні на заводі немає, виходить, черга механізмів перед майстернею може рости практично необмежено. Потрібно обчислити імовірності станів системи; середнє число заявок в черзі на обслуговування; середнє число заявок, що перебувають в системі; середню тривалість перебування заявки в черзі; середню тривалість перебування заявки в системі.

Розв’язання: Визначимо параметр потоку обслуговувань:

$$\mu = \frac{1}{\bar{t}_{об}} = \frac{1}{0,5} = 2 .$$

Приведена інтенсивність потоку заявок

$$\alpha = \frac{2,5}{2} = 1,25 \ ,$$

при цьому $$\frac{\alpha}{S} = \frac{2{,}5}{2*3} = 0{,}41 .$$

Оскільки $\frac{\alpha}{S} < 1$, то черга не росте безмежно, і в системі наступає граничний стаціонарний режим роботи.

Обчислимо імовірності станів системи:

$$p_0 = \left[\sum_{n=0}^{S-1} \frac{\alpha^n}{n!} + \frac{\alpha^S}{S!\left(1 - \frac{\alpha}{S}\right)} \right]^{-1} = \frac{1}{1 + \frac{\alpha^1}{1!} + \frac{\alpha^2}{2!} + \frac{\alpha^3}{3!*\left(1 - \frac{\alpha}{3}\right)}} =$$

$$= \frac{1}{1 + \alpha + \frac{\alpha^2}{2!} + \frac{\alpha^3}{6*\left(1 - \frac{\alpha}{3}\right)}} = \frac{1}{1 + 1{,}25 + \frac{(1{,}25)^2}{2!} + \frac{(1{,}25)^3}{6*\left(1 - \frac{1{,}25}{3}\right)}} = 0{,}279;$$

$$p_1 = \frac{\alpha^1}{1!} p_0 = 1{,}25 * 0{,}279 = 0{,}349$$

$$p_2 = \frac{\alpha^2}{2!} p_0 = \frac{(1{,}25)^2}{2} * 0{,}279 = 0{,}218$$

$$p_3 = \frac{\alpha^3}{3!} P_0 = \frac{(1{,}25)^3}{6} * 0{,}279 = 0{,}091$$

$$p_4 = \frac{\alpha^4}{4!} p_0 = \frac{(1{,}25)^4}{24} * 0{,}279 = 0{,}028 .$$

Імовірність відсутності черги в майстерні

$$P_{відс} = p_0 + p_1 + p_2 + p_3 = 0{,}279 + 0{,}349 + 0{,}218 + 0{,}091 = 0{,}937 .$$

Середнє число заявок в черзі на обслуговування

$$L_q = \frac{S\alpha}{(S - \alpha)^2} p_s = \frac{3 * 1{,}25}{(3 - 1{,}25)^2} 0{,}091 = 0{,}111 .$$

Середнє число заявок, що перебувають в системі

$$L_s = L_q + \alpha = 0{,}111 + 1{,}25 = 1{,}361 .$$

Середня тривалість перебування механізму в черзі на обслуговування

$$W_q = \frac{L_q}{\lambda} = \frac{1{,}361}{2{,}5} = 0{,}044 \text{ доби.}$$

Середня тривалість перебування механізму в майстерні (в системі)

$$W_s = W_q + \frac{1}{\mu} = 0{,}044 + \frac{1}{2} = 0{,}544 \text{ доби.}$$

Комп'ютерна обробка статистичних даних

Обробка даних – це процес аналітичного дослідження великих масивів інформації (звичайно економічного характеру) з метою виявлення певних закономірностей і систематичних взаємозв'язків між змінними, які потім можна ужити до нових сукупностей даних. Цей процес включає три основних етапи:

дослідження, побудова моделі або структури і її перевірка. В реальній ситуації практично неможливо перевірити економічну модель на стадії аналізу, тому початкові результати мають характер евристик, які можна використовувати в процесі прийняття рішення. На противагу алгоритму, який описує цілком певний набір операцій для одержання конкретного результату, евристики - це загальні рекомендації або поради, засновані на статистичній очевидності.

Для аналізу статистичних даних звичайно використовують пакети прикладних статистичних програм, в числі яких, крім Microsoft Excel, можна назвати такі як STATISTICA, STATGRAF, SPSS, SAS і ін.

Для систем, в яких бере участь велика кількість різнорідних елементів, а випадкові фактори складним способом взаємодіють між собою, процеси, що протікають, неможливо описати аналітично як за допомогою детермінованих, так і імовірнісних методів. У цих випадках використовують метод імітаційного моделювання, що, як правило, виявляється простіше аналітичного, а нерідко і єдино можливим методом. Зокрема, основним допущенням, при якому аналізувалися розглянуті СМО, є те, що всі потоки, які переводять їх з одного стану в інший, були найпростішими. При порушенні цієї вимоги загальних аналітичних методів для таких систем дістати не удається. У цьому разі і користуються універсальним методом статистичних випробувань або методом імітаційного моделювання.

У першому наближенні ідею методу можна описати так: процес функціонування складної системи імітується на ЕОМ з усіма випадками, що його супроводжують. Таким чином, будується одна реалізація випадкового процесу з випадковим ходом і результатом. Сама по собі одна реалізація не дає підстав до вибору рішення, але, діставши множину таких реалізацій, можна їх обробити як звичайний статистичний матеріал, визначити середні характеристики за множиною і дістати уявлення про те, які умови задачі й елементи рішення впливають на функціонування системи. Таким чином, при використанні методу імітаційного моделювання випадковість використовується як апарат дослідження.

Методи обробки даних здобувають все більшу популярність як інструмент для аналізу економічної інформації, особливо в тих випадках, коли передбачається, що з наявних даних можна буде одержати знання для прийняття рішень в умовах невизначеності. Хоча останнім часом зріс інтерес до розробки нових методів аналізу даних, спеціально призначених для сфери бізнесу, в цілому системи обробки даних, як і раніше, ґрунтуються на класичних принципах аналізу даних і побудови моделей і використовують ті ж підходи й методи.

Іноді, при проведенні аналізу лінійної моделі, дослідник отримує дані про те, що модель неадекватна. У цьому випадку його, як і раніше, цікавить залежність між незалежними і залежною змінними, але для уточнення моделі в рівняння додаються деякі нелінійні члени. Наприклад, його можна використовувати для уточнення залежності між стажем, роботи і продуктивністю праці, вартістю будинку і часом, необхідним для його продажу і т.д. Подібні задачі часто вирішуються методами множинної регресії і дисперсійного аналізу. Однак і в методі множинної регресії, і в дисперсійному аналізі припускається, що залежність між незалежними і залежною змінними лінійна. При нелінійному оцінюванні вибір характеру залежності залишається за дослідником. Наприклад, можна визначити залежну змінну як логарифмічну функцію від незалежної змінної, як степеневу функцію або як будь-яку іншу композицію елементарних функцій від незалежних змінних.

В останні роки в практику розв'язання задач прогнозування і керування ввійшли нейронні мережі. Вони успішно застосовуються у різних областях - бізнесі, медицині, техніці, геології, фізиці. Нейронні мережі - винятково могутній метод моделювання, що дозволяє відтворювати надзвичайно складні залежності. Зокрема, нейронні мережі легко реалізують нелінійні залежності. Лінійне моделювання протягом багатьох років було основним методом моделювання в більшості задач, оскільки для нього добре розроблений математичний апарат. У задачах же з великим числом змінних моделювати лінійні залежності дуже складно, і нейронні мережі справляються з цією проблемою.

Нейронні мережі - це клас аналітичних методів, заснований на ідеї відтворення процесів навчання мислячих істот і функцій нервових кліток. Нейронні мережі можуть прогнозувати майбутні значення змінних за вже наявними значеннями цих же або інших змінних, попередньо здійснивши процес так званого навчання на основі наявних даних.

Навчання нейронних мереж здійснюється на прикладах. Користувач нейронної мережі підбирає необхідні дані, а потім включає алгоритм навчання, що автоматично сприймає структуру даних. При цьому від користувача, звичайно, потрібний якийсь набір евристичних знань про те, як варто відбирати і підготовляти дані, вибирати потрібну архітектуру мережі й інтерпретувати результати, однак рівень знань, необхідний для успішного застосування нейронних мереж, набагато скромніше, ніж, наприклад, при використанні традиційних методів математичної статистики.

Нейронні мережі засновані на примітивній біологічній моделі нервових систем. Вони виникли з досліджень в області штучного інтелекту при спробах відтворити здатність біологічних нервових систем навчатися і виправляти помилки, моделюючи низькорівневу структуру мозку. Нейрони - це спеціальні клітини, здатні розповсюджувати електрохімічні сигнали. Нейрон має розгалужену структуру введення інформації (дендрити), ядро і вихід, що розгалужується (аксон). Аксони клітини з'єднуються з дендритами інших кліток за допомогою синапсів. При активації нейрон посилає електрохімічний сигнал по своєму аксонові. Через синапси цей сигнал досягає інших нейронів, які можуть в свою чергу активуватися. Нейрон активується тоді, коли сумарний рівень сигналів, що прийшли в його ядро з дендритів, перевищить певний рівень (поріг активації). Отже, будучи побудований з дуже великого числа зовсім простих елементів, кожний з яких приймає зважену суму вхідних сигналів, і у випадку, якщо сумарний вхід перевищує певний рівень, передає далі двійковий сигнал, мозок здатен вирішувати надзвичайно складні завдання.

Базова штучна модель нейронної мережі побудована в такий спосіб. Штучний нейрон отримує вхідні сигнали (вихідні дані або вихідні сигнали інших нейронів) через кілька вхідних каналів. Кожен вхідний сигнал проходить через з'єднання, що має певну інтенсивність (або вагу), ця вага відповідає синаптичної активності біологічного нейрона. З кожним нейроном пов'язане певне граничне значення. Обчислюється зважена сума входів, з неї віднімається граничне значення, в результаті утворюється величина активації нейрона.

Сигнал активації перетворюється за допомогою функції активації (передаточної функції) і в результаті утворюється вихідний сигнал нейрона. Якщо при цьому використовується східчаста функція активації (тобто вихід нейрона дорівнює нулю, якщо вхід від'ємний, і одиниці, якщо вхід нульовий або додатний), то такий нейрон працюватиме точно так само, як природний нейрон (відняти граничне значення зі зваженої суми і порівняти результат з нулем - це те ж саме, що порівняти зважену суму з граничним значенням). В дійсності, східчасті функції рідко використовуються в штучних нейронних мережах. Помітимо також, що ваги можуть бути від'ємними, - це означає, що синапс надає нейронові не збуджуючий, а гальмуючий вплив (у мозку присутні гальмуючі нейрони).

Таким чином, мережа має входи, що приймають значення змінних із зовнішнього миру, і виходи - прогнози або керуючі сигнали. Найпростіша мережа має структуру прямої передачі сигналу. Сигнали проходять від входів через приховані елементи і в результаті приходять на вихідні елементи. Така структура має стале поводження. Іноді використовують рекурентні мережі, в яких містяться зв'язки, що ведуть назад від більш далеких до більш ближніх нейронів. Така мережа може бути нестійка і мати дуже складну динаміку поводження. Однак при розв'язанні практичних задач найбільш корисними виявилися структури прямої передачі.

Мережа з прямою передачею сигналу показана на рис. 11.11. Нейрони певним чином організовані в шари. Вхідний шар слугує просто для введення значень вхідних змінних. Кожен з прихованих і вихідних нейронів з'єднаний з усіма елементами попереднього шару. Можна було б розглядати мережі, в яких нейрони пов'язані тільки з деякими з нейронів попереднього шару, однак, для більшості додатків мережі з повною системою сполучень переважніше.

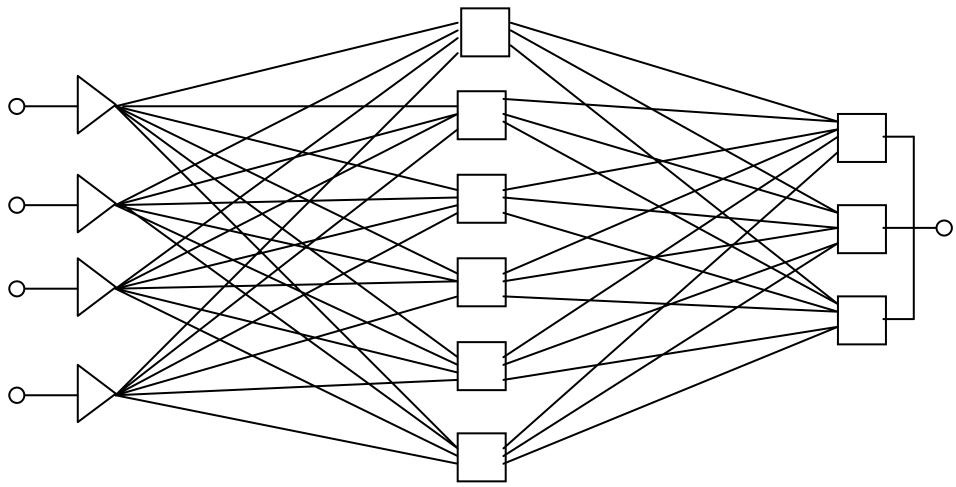


Рис. 11.11 - Структура нейронної мережі

При використанні мережі на вхідні елементи подаються значення вхідних змінних, потім послідовно відпрацьовують нейрони проміжних і вихідних шарів. Кожен з них обчислює своє значення активації, беручи зважену суму виходів елементів попереднього шару і віднімаючи з неї граничне значення. Потім значення активації перетворяюються за допомогою функції активації і в результаті утворюється вихід нейрона. Вихідні значення елементів вихідного шару приймаються за вихід всієї мережі в цілому.

Клас задач, які можна вирішити за допомогою нейронної мережі, визначається тим, як мережа працює і тим, як вона навчається. Нейронна мережа приймає значення вхідних змінних і видає значення вихідних змінних. Отже, мережу можна застосовувати в ситуації, коли є певна відома інформація і з неї необхідно дістати якусь поки не відому інформацію. Приклади таких задач: 1) Прогнозування на фондовому ринку. Знаючи ціни акцій за останній тиждень і сьогоднішнє значення індексу FTSE, необхідно спрогнозувати завтрашню ціну акцій. 2) Надання кредиту. Потрібно визначити, чи високий ризик надання кредиту приватній особі, що звернулася з таким проханням. В результаті розмови з ним відомий його прибуток, що попередня кредитна історія і т.д. 3) Керування. Треба визначити, що повинен робити робот (повернутися праворуч або ліворуч, рухатися вперед і т.д.), щоб досягти мети; відоме зображення, що передає встановлена на роботі відеокамера.

Зрозуміло, не будь-яку задачу можна вирішити за допомогою нейронної мережі. Необхідно, щоб між відомими вхідними значеннями і невідомими виходами був зв'язок. Цей зв'язок може бути перекручений впливом випадкових факторів (зокрема, неможливо за даними з приклада із прогнозуванням цін акцій побудувати абсолютно точний прогноз, оскільки на ціну впливають і інші фактори, не представлені у вхідному наборі даних, і крім того в задачі присутній елемент випадковості), але вона повинна існувати.

Як правило, нейронна мережа використовується тоді, коли невідомий точний вигляд зв'язків між входами і виходами, - якби він був відомий, то зв'язок можна було б моделювати безпосередньо. Інша істотна особливість нейронних мереж полягає в тому, що залежність між входом і виходом перебуває в процесі навчання мережі. Якщо мережа навчена добре, вона здобуває здатність моделювати невідому функцію, що зв'язує значення вхідної і вихідної змінних, і згодом таку мережу можна використовувати для прогнозування в ситуації, коли вихідні значення невідомі.

Нейронні мережі можуть працювати з числовими даними, що лежать у певному обмеженому діапазоні. Більш важким завданням є робота з даними нечислового характеру. Найчастіше нечислові дані бувають представлені у вигляді номінальних змінних типу Стать = {Чол., Жін.}. Змінні з номінальними значеннями можна уявити в числовому вигляді. Нехай, наприклад, необхідно навчити нейронну мережу оцінювати вартість об'єктів нерухомості. Ціна будинку сильно залежить від того, в якому районі міста він розташований. Місто може бути поділене на кілька десятків районів, що мають власні назви, і здається природним запровадити для позначення району змінну з номінальними значеннями. На жаль, в цьому випадку навчити нейронну мережу буде дуже важко, замість цього краще присвоїти кожному районові певний рейтинг, ґрунтуючись на експертних оцінках.

У багатьох реальних задачах доводиться мати справу з не зовсім достовірними даними. Значення деяких змінних можуть бути перекручені або частково відсутні.

Усяка нейронна мережа приймає на вході числові значення і видає на виході також числові значення. Передаточна функція для кожного елемента мережі звичайно вибирається таким чином, щоб її вхідний аргумент міг приймати довільні значення, а вихідні значення лежали б у строго обмеженому діапазоні. При цьому хоча вхідні значення можуть бути будь-якими, виникає ефект насичення, коли елемент виявляється чутливим лише до вхідних значень, що лежать в деякій обмеженій області. На рис. 11.12 представлена одна з найпоширеніших передаточних функцій - так звана логістична функція (іноді її називають сигмоїдною функцією, тому що вона має форму літери S). У цьому випадку вихідне значення завжди буде лежати в інтервалі (0,1), а область чутливості для входів трохи ширше інтервалу (-1,+1). Дана функція є гладкою, а її похідна легко обчислюється - це обставина дуже істотна для роботи алгоритму навчання мережі (в цьому також криється причина того, що східчаста функція для цієї мети практично не використовується).

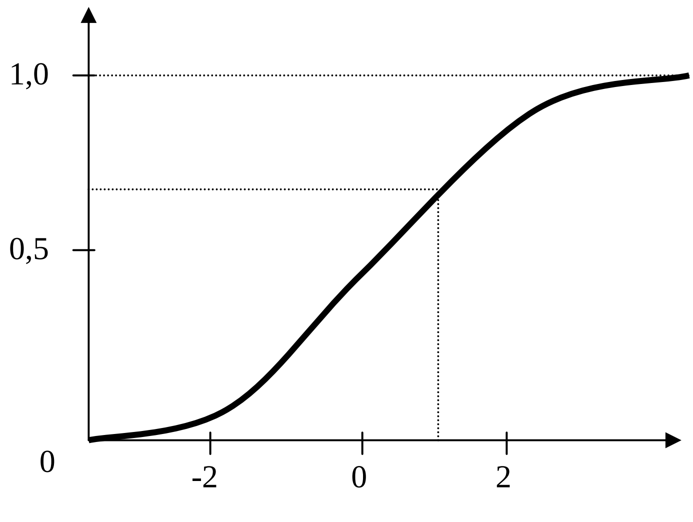


Рис. 11.12 - Крива логістичної функції

Оскільки вихідні значення завжди належать деякій обмеженій області, а вся інформація повинна бути подана в числовому вигляді, очевидно, що при розв'язанні реальних задач методами нейронних мереж потрібні етапи попередньої обробки - пре-процесування - і заключної обробки - пост-процесування даних. Етапи пре-процесування і пост-процесування даних дозволяють привести сирі вихідні дані в числову форму, придатну для обробки нейронною мережею, і перетворити вихід нейронної мережі назад у формат вхідних даних. Нейронна мережа служить "прошарком" між пре- і пост- процесуванням, і результат видається в потрібному вигляді (наприклад, в задачі класифікації видається назва вихідного класу).

Найбільш часто використовується архітектура мережі, називана «багатошаровим персептроном». Кожен елемент мережі будує зважену суму своїх входів з виправленням у вигляді доданка і потім пропускає цю величину активації через передаточну функцію, в такий спосіб утворюється вихідне значення цього елемента. Елементи організовані в пошарову топологію з прямою передачею сигналу. Таку мережу легко можна інтерпретувати як модель вхід-вихід, в якій ваги й граничні значення (зсуви) є вільними параметрами моделі. Така мережа може моделювати функцію практично будь-якого ступеня складності, причому число шарів і число елементів в кожному шарі визначають складність функції. Після того, як визначене число шарів і число елементів в кожному з них, потрібно знайти значення для ваг і порогів мережі, які б мінімізували помилку прогнозу, видаваного мережею. Саме для цього служать алгоритми навчання. З використанням зібраних історичних даних ваги і граничні значення автоматично коректуються з метою мінімізувати помилку. По суті, цей процес являє собою припасування моделі, що реалізується мережею, до наявних навчальних даних. Помилка для конкретної конфігурації мережі визначається шляхом прогону через мережу всіх наявних спостережень і порівняння реально видаваних вихідних значень із бажаними (цільовими) значеннями. Всі такі різниці підсумовуються в так звану функцію помилок, значення якої і є помилка мережі. Як функцію помилок найчастіше беруть суму квадратів помилок.

Головним показником якості результату є контрольна помилка. При цьому відповідно до загальнонаукового принципу, згідно з яким при інших рівних слід віддати перевагу більш простій моделі, з двох мереж з приблизно рівними помилками контролю має сенс вибрати ту, котра менше.

Таким чином, побудова мережі складається з наступних кроків:

- вибрати початкову конфігурацію мережі (наприклад, один проміжний шар з числом елементів у ньому, рівним напівсумі числа входів і числа виходів;
- провести ряд експериментів з різними конфігураціями, запам'ятовуючи при цьому кращу мережу (в смислі контрольної помилки);
- якщо в черговому експерименті спостерігається недонавчання (мережа не видає результат прийнятної якості), додати додаткові нейрони в проміжний шар. Якщо це не допомагає, спробувати додати новий проміжний шар.

На всіх етапах істотною є вимога, щоб навчальна, контрольна і тестова множини були репрезентативними з погляду суті задачі (більше того, ці множини повинні бути репрезентативними кожна окремо). Якщо навчальні дані не репрезентативні, то модель, як мінімум, буде не дуже гарною, а в найгіршому разі - марною.

Теоретично, для моделювання будь-якої задачі досить багатошарового персептрона з двома проміжними шарами (в точному формулюванні цей результат відомий як теорема Колмогорова). При цьому може бути і так, що для розв'язання деякої конкретної задачі більш простою і зручною буде мережа з ще більшим числом шарів. Однак для розв'язання більшості практичних задач досить одного проміжного шару, два шари застосовують як резерв в особливих випадках, а мережі з трьома шарами практично не використовують.

Задачі прогнозування можна розбити на два основних класи - задачі класифікації і задачі регресії.

Класифікація - це віднесення спостереження до одного з кількох, заздалегідь відомих класів (представлених значеннями номінальної вихідної змінної). У задачах класифікації потрібно визначити, до якого з декількох заданих класів належить даний вхідний набір. Прикладами можуть слугувати надання кредиту (чи відноситься дана особа до групи високого або низького кредитного ризику), розпізнавання підпису (підроблений, справжній) та ін. У цих випадках на виході повинна бути всього одна номінальна змінна.

До регресії відносять категорію задач, де мета полягає в тому, щоб оцінити значення безперервної вихідної змінної за значеннями вхідних змінних. У задачах регресії потрібно пророчити значення змінної, що приймає, як правило, безперервні числові значення: завтрашню ціну акцій, витрату палива в автомобілі, прибутку в наступному році і т.п. У таких випадках в якості вихідної повинна бути одна числова змінна.

Нейронна мережа може вирішувати одночасно кілька задач регресії і класифікації, однак звичайно в кожний момент часу вирішується тільки одна задача. Таким чином, в більшості випадків нейронна мережа матиме всього одну вихідну змінну. У випадку розв'язання задач класифікації з багатьма станами може знадобитися кілька вихідних елементів.

У задачах класифікації вихідний елемент повинен видавати сильний сигнал у випадку, якщо дане спостереження належить до класу, що цікавить нас, і слабкий - у протилежному разі. Інакше кажучи, цей елемент повинен прагнути змоделювати функцію, рівну одиниці в тій області простору об'єктів, де розташовуються об'єкти з потрібного класу, і рівну нулю поза цією областю. Така конструкція відома як дискримінантна функція в задачах розпізнавання.

У задачах класифікації інтерпретацію точок, що лежать між глобальним максимумом і глобальним мінімумом знаходять у такий спосіб. Для граничних значень встановлюють деякі довірчі межі (прийняття або відкидання), які повинні бути досягнуті, щоб даний елемент вважався "прийнявшим рішення". Наприклад, якщо встановлені пороги прийняття/відкидання 0,95/0,05, то при рівні вихідного сигналу, що перевершує 0,95, елемент вважається активним, при рівні нижче 0,05 - неактивним, а в проміжку - "невизначеним". Є і більш тонкий спосіб інтерпретувати рівні вихідного сигналу: вважати їх імовірностями. У цьому випадку мережа видає трохи більшу інформацію, ніж просто "так/ні": вона повідомляє нам, наскільки ми можемо довіряти її рішенню. Розроблено модифікації методу „Багатошаровий персептрон”, що дозволяють інтерпретувати вихідний сигнал нейронної мережі як імовірність, в результаті чого мережа, по суті, вчиться моделювати щільність імовірності розподілу даного класу. При цьому, однак, імовірнісна інтерпретація обґрунтована тільки в тому випадку, якщо виконані певні припущення щодо розподілу вихідних даних. Тут, як і раніше, може бути ухвалене рішення щодо класифікації, але, крім того, імовірнісна інтерпретація дозволяє запровадити концепцію "розв'язання з мінімальними витратами".

Таким чином, говорячи про задачі класифікації, ми вказали, що виходи мережі можна інтерпретувати як оцінки імовірності того, що елемент належить деякому класу, і мережа фактично вчиться оцінювати функцію щільності імовірності. Аналогічна інтерпретація може мати місце і в задачах регресії. Вихід мережі розглядається як очікуване значення моделі в даній точці простору входів. Це очікуване значення пов'язане з щільністю імовірності спільного розподілу вхідних і вихідних даних.

Задача оцінки щільності імовірності за даними має давню історію в математичній статистиці і належить до області бейесової статистики. Звичайна статистика за заданою моделлю говорить, яка буде імовірність того або іншого результату (наприклад, що на гральній кістці шість очок буде випадати в середньому в одному випадку з шістьох). Бейесова статистика дозволяє оцінити правильність моделі за наявними достовірними даними. У більш загальному плані бейесова статистика дає можливість оцінювати щільність імовірності розподілів параметрів моделі за наявними даними. Для того щоб мінімізувати помилку, вибирається модель з такими параметрами, при яких щільність імовірності буде найбільшою.

При розв'язанні задачі класифікації можна оцінити щільність імовірності для кожного класу, порівняти між собою імовірності належності різним класам і вибрати найбільш імовірний. Саме це відбувається при навчанні нейронної мережі розв'язанню задачі класифікації - мережа намагається визначити (тобто апроксимувати) щільність імовірності.

Традиційний підхід до задачі полягає в тому, щоб побудувати оцінку для щільності імовірності за наявними даними. Звичайно при цьому припускається, що щільність має деякий певний вигляд, найчастіше що вона має нормальний розподіл. Після цього оцінюються параметри моделі. Нормальний розподіл часто використовується тому, що в цьому випадку параметри моделі (середнє і стандартне відхилення) можна оцінити аналітично.

Таким чином, ми прийшли до понять імовірнісної нейронної мережі і узагальнено-регресійної нейронної мережі. Імовірнісні нейронні мережі призначені для задач класифікації, а узагальнено-регресійні нейронні мережі - для задач регресії.

Імовірнісна нейронна мережа має єдиний керуючий параметр навчання, значення якого повинне вибиратися користувачем, - ступінь згладжування (або відхилення гауссової функції). Найбільш важливі переваги імовірнісних нейронних мереж полягають в тому, що вихідне значення має імовірнісний сенс (і тому його легше інтерпретувати), і в тому, що мережа швидко навчається. Істотним недоліком таких мереж є їхній об'єм. Імовірнісна нейронна мережа фактично вміщує в себе всі навчальні дані, тому вона вимагає багато пам'яті і може повільно працювати.

Узагальнено-регресійна нейронна мережа влаштована аналогічно імовірнісній нейронній мережі, але вона призначена для розв'язання задач регресії, а не класифікації. Вважається, що кожне спостереження свідчить про деяку впевненість в тому, що поверхня відгуку в даній точці має певну висоту, і ця впевненість убуває при відході убік від точки. Узагальнено-регресійна нейронна мережа копіює усередину себе всі навчальні спостереження і використовує їх для оцінки відгуку в довільній точці. Остаточна вихідна оцінка мережі виходить як зважене середнє виходів за всіма навчальними спостереженнями, де величини ваг відображають відстань від цих спостережень до тієї точки, в якій проводиться оцінювання. Таким чином, більш близькі точки вносять більший вклад в оцінку.

Відповідно до загальноприйнятого в науці принципу, якщо більш складна модель не дає кращих результатів, ніж більш проста, то з них слід віддати перевагу другій. У термінах апроксимації відображень найпростішою моделлю буде лінійна, в якій підгінна функція визначається гіперплощиною. У задачі класифікації гіперплощина розміщується таким чином, щоб вона розділяла собою два класи (лінійна дискримінантна функція); в задачі регресії гіперплощина повинна проходити через задані точки.

Мовою нейронних мереж лінійна модель представляється мережею без проміжних шарів, що у вихідному шарі містить тільки лінійні елементи, тобто елементи з лінійною функцією активації. Ваги відповідають елементам матриці, а пороги - компонентам вектора зсуву. Під час роботи мережа фактично множить вектор входів на матрицю ваг, а потім до отриманого вектора додає вектор зсуву.

Мережі Кохонена принципово відрізняються від всіх інших типів мереж. У той час як всі інші мережі призначені для задач з керованим навчанням, мережа Кохонена головним чином розрахована на некероване навчання.

При керованому навчанні спостереження, що складають навчальні дані, разом з вхідними змінними містять також і відповідні їм вихідні значення, і мережа повинна відновити відображення, що переводить перші в другі. У випадку ж некерованого навчання навчальні дані містять тільки значення вхідних змінних. На перший погляд це може здатися дивним. Як мережа зможе чомусь навчитися, не маючи вихідних значень? Відповідь полягає в тому, що мережа Кохонена вчиться розуміти саму структуру даних.

Одне з можливих застосувань таких мереж - розвідницький аналіз даних. Мережа Кохонена може розпізнавати кластери в даних, а також встановлювати близькість класів. Таким чином, користувач може поліпшити своє розуміння структури даних, щоб потім уточнити нейромережну модель. Якщо в даних розпізнані класи, то їх можна позначити, після чого мережа зможе вирішувати задачі класифікації. Мережі Кохонена можна використовувати і в тих задачах класифікації, де класи вже задані, - тоді перевага буде в тому, що мережа зможе виявити подібність між різними класами. Інша можлива область застосування - виявлення нових явищ. Мережа Кохонена розпізнає кластери в навчальних даних і відносить всі дані до тих або інших кластерів. Якщо після цього мережа зустрінеться з набором даних, несхожим ні на один з відомих зразків, то вона не зможе класифікувати такий набір і тим самим виявить його новизну.

Мережа Кохонена має всього два шари: вхідний і вихідний, складений з радіальних елементів (вихідний шар називають ще шаром топологічної карти). Елементи топологічної карти розташовуються в деякому просторі - як правило, двовимірному.

Крім того, що вже сказано, в результаті ітеративної процедури навчання мережа організується таким чином, що елементи, які відповідають центрам, розташованим близько один від одного в просторі входів, будуть розташовуватися близько один від одного і на топологічній карті. Топологічний шар мережі можна уявляти собі як двовимірну решітку, яку потрібно так відобразити в N-мірному просторі входів, щоб по можливості зберегти вихідну структуру даних. Звичайно, при будь-якій спробі уявити N-мірний простір на площині будуть загублені багато деталей; однак такий прийом іноді корисний, тому що він дозволяє користувачеві візуалізовати дані, які ніяким іншим способом зрозуміти неможливо.

У задачах регресії метою є оцінка значення числової змінної, приймаючої безперервний діапазон значень, вихідної змінної за значеннями вхідних змінних. Для розв'язання задач регресії можна використовувати мережі наступних типів: багатошаровий персептрон, узагальнено-регресійну мережу і лінійну мережу. При цьому вихідні дані повинні мати стандартний числовий (не номінальний) тип.

Особливу важливість для регресії мають масштабування вихідних значень і ефекти екстраполяції (прогнозування невідомих значень шляхом продовження функцій за межі області відомих значень). Нейронні мережі найбільше часто використовуваних архітектур видають вихідні значення в деякому певному діапазоні (наприклад, на відрізку [0,1] у випадку логістичної функції активації). Для задач класифікації це не створює труднощів. Однак для задач регресії тут є проблема. Якість роботи мережі в задачі регресії можна перевірити декількома способами. По-перше, мережі можна повідомити вихідне значення, що відповідає будь-якому спостереженню (або якомусь новому спостереженню, яке необхідно перевірити). Якщо це спостереження містилося у вихідних даних, то видається значення різниці. По-друге, можуть бути отримані підсумкові статистики. До них належать середнє значення і стандартне відхилення, обчислені для навчальних даних і для помилки прогнозу. В загальному випадку середнє значення помилки прогнозу буде дуже близько до нуля (зрештою, нульове середнє для помилки прогнозу можна дістати, попросту оцінивши середнє значення навчальних даних і зовсім не звертаючись до значень вхідних змінних). Найбільш важливим показником якості моделі є стандартне відхилення помилки прогнозу. Якщо воно не виявиться істотно менше стандартного відхилення навчальних даних, це означатиме, що мережа працює не краще, ніж проста оцінка за середнім. Використовується також відношення стандартного відхилення помилки прогнозу до стандартного відхилення навчальних даних. Якщо воно істотно менше одиниці (наприклад, нижче 0,1), то це свідчить про гарну якість регресії. Величину, отриману як різниця між одиницею і цим регресійним відношенням, іноді називають часткою поясненої дисперсії моделі.

Запитання для самоперевірки:

1. Дайте визначення випадкового процесу.
2. Що таке реалізація випадкової функції.
3. Які властивості імовірнісних характеристик стаціонарного випадкового процесу?
4. Дайте визначення кореляційної функції.
5. Поясніть властивість ергодичності стаціонарного випадкового процесу.
6. Які задачі вирішує теорія масового обслуговування?
7. Перелічіть основні компоненти системи масового обслуговування, що обумовлюють її математичний опис.
8. Поясніть класифікацію систем масового обслуговування.
9. В якому випадку випадковий процес, що протікає в системі, називається марковським?
10. Дайте визначення стаціонарного потоку подій і регулярного потоку подій.
11. Які змінні є невідомими в рівняннях Колмогорова? Що таке фінальної імовірності?
12. Якими параметрами визначається потік заявок? Потік обслуговувань?
13. В чому полягає суть методу статистичних випробувань?
14. В чому полягає суть моделювання в нейронних мережах?

Задачі для самостійного розв'язання

9.1. Випадкова функція X(t) в кожному перетині являє собою безперервну випадкову величину з щільністю розподілу f(x,t). Напишіть вираз для математичного сподівання $m_x(t)$ і дисперсії $D_x(t)$ випадкової функції X(t).

9.2. Випадкова функція X(t) задана у вигляді X(t) = Vt+1, де V – випадкова величина, розподілена за нормальним законом з параметрами m_v=1 і σ_v=0,2. Визначити числові характеристики випадкової функції X(t).

9.3. Розглядається робота автоматичної телефонної станції, розрахованої на одночасне обслуговування 20 абонентів. Виклик на АТС надходить в середньому через 6 секунд. Кожна розмова триває в середньому 2 хвилини. Якщо абонент застає АТС зайнятою, він отримує відмову. Якщо абонент застає вільним хоча б один з 20 каналів, він з'єднується з потрібним йому номером. Визначити імовірність того, що абонент буде обслужений, а також інші характеристики роботи СМО: середнє число зайнятих каналів, імовірність зайнятості каналу, середній час простою каналу.

РОЗДІЛ 3
ЗАВДАННЯ ДЛЯ САМОСТІЙНОЇ РОБОТИ

При самостійній роботі над матеріалом курсу „Теорія імовірностей і математична статистика" необхідно вивчити теоретичні положення, вивчити визначення, розібрати приклади вирішених задач. З метою більш глибокого вивчення матеріалу рекомендується звертатися до основної і додаткової літератури, наведеної наприкінці посібника. При самостійній роботі над матеріалом курсу треба вести конспект і термінологічний словник. Після освоєння певної теми необхідно відповісти на запитання для самоперевірки і розв'язати задачі, пропоновані для самостійного вирішення.

Після вивчення кожного змістового модуля необхідно підготуватися до поточного тестування, відповіді на тести слід навести в зошиті для практичних занять.

Нижче наведено тести для поточного контролю знань, які розташовані за змістовими модулями і темами. Далі йде завдання для виконання контрольної роботи з курсу і запитання, на які треба відповісти при здачі заліку.

ТЕСТОВІ ЗАВДАННЯ З КУРСУ „ТЕОРІЯ ІМОВІРНОСТЕЙ І МАТЕМАТИЧНА СТАТИСТИКА”

ТЕСТИ до ЗМ1. ТЕОРІЯ ІМОВІРНОСТЕЙ

7

Завдання № 1

Визначте відповідність у вигляді комбінації цифр і літер:

1) Явище при якому повторення досліду незмінно викликає певну подію;	А) Неможлива подія
2) Випадкове явище, що містить статистичну однорідність;	Б) Дослід
3) Всякий факт, що у результаті досліду може відбутися або не відбутися;	В) Стохастичне
4) Відтворена сукупність умов, у яких може відбутися випадкова подія;	Г) Випадкова подія
5) Числова міра ступеня об'єктивної можливості появи випадкової події в результаті досліду;	Д) Детерміноване
6) Подія, що в результаті досліду обов'язково повинна відбутися;	Е) Достовірна подія
7) Подія, що в результаті даного досліду ніяк відбутися не може.	Ж) Імовірність випадкової події

1._____2._____3._____4._____5._____6._____7.______

6

Завдання № 2

Визначте відповідність у вигляді комбінації цифр і літер.

1) Математичне сподівання;	А) Me
2) Мода;	Б) σ
3) Медіана;	В) Mx
4) Імовірність;	Г) P(A), p, q
5) Середнє квадратичне відхилення;	Д) Mo
6) Дисперсія.	Е) Dx

1.______2.______3.______4.______5.______6._______

1

Завдання № 3

Доповніть твердження, вписавши одне слово:

За одиницю виміру імовірності приймають імовірність ____події.

5

Завдання № 4

Визначте відповідність у вигляді комбінації цифр і літер.

1) випадкова величина;

2) випадкова подія;

А) Число народжених хлопчиків серед 100 немовлят;

Б) Поява двох гербів при трьох підкиданнях монети;

В) Витяг з урни червоної кулі;

Г) Відстань, що пролетить снаряд при вильоті зі знаряддя;

Д) Поразка мішені з одного пострілу.

1._____________________________ 2._________________________________

4

Завдання № 5

Доповніть твердження, вписавши кілька словосполучень:

Імовірність випадкової події можна визначити класичним методом, якщо наслідки досліду мають наступні властивості:

1.____

2.____

3.____

1

Завдання № 6

Доповніть твердження, вписавши одне слово:

Відношення числа сприяючих події А наслідків до загального числа елементарних наслідків досліду називають __________________________ події А.

2

Завдання № 7

Вставте в дужки відповідні цифри.

Імовірність випадкової події перебуває в межах $[\ ]\le P(A)\le[\ ]$.

1

Завдання № 8

Доповніть ствердження, вписавши одне слово:

Якщо кожний елементарний результат випробування сприяє події, то таку подію називають ______________________________ .

1

Завдання № 9

Доповніть ствердження, вписавши одне слово:

Якщо поява однієї з декількох подій виключає появу інших подій у тому самому випробуванні, то такі події називають

____

1

Завдання № 10

Доповніть ствердження, вписавши одне слово:

Якщо є підстави вважати, що одна з ряду подій не є більш можливою, ніж кожна з останніх, то такі події називають

______________________________.

3

Завдання № 11

Вставте в скобки відповідні цифри.

Імовірність достовірної події дорівнює $\frac{[\ ]}{n}=[\ ]$

3

Завдання № 12

Вставте в скобки відповідні цифри.

Імовірність неможливої події дорівнює $\frac{[\ ]}{n}=[\ ]$

4

Завдання № 13

Розвяжіть задачу:

З партії М деталей, що містить n бракованих, для контролю навмання виймають K деталей. Визначити імовірність того, що в числі деталей, взятих для контролю виявиться d бракованих.

$P(A)=$ ______________

2

Завдання № 14

Вибрати правильні відповіді.

Монета підкинута п'ять разів. Подія А полягає в появі хоча б двох решок.

1) дві решки;
2) два герби;
3) дві або більше решки;
4) нуль або одна решка;
5) дві або три або чотири решки;
6) дві або три або чотири або п'ять решок.

____

1

Завдання № 15

Доповніть твердження, вписавши одне слово:

Подія, що перебуває в появі події А або події В, або обох цих подій, називають ______________________________ двох подій.

1

Завдання № 16

Доповніть твердження, вписавши одне слово:

Подія, що перебуває в появі події А и події В одночасно, називають ______________________________ двох подій.

1

Завдання № 17

Доповніть твердження, вписавши одне слово:

Дві несумісних події А і $\bar{A}$, що складають повну групу, називаються

______________________________ .

1

Завдання № 18

Вставте в дужки відповідну цифру.

Сума імовірностей протилежних подій дорівнює $P(A)+P(\overline{A})=[\ ]$.

1

Завдання № 19

Доповніть твердження, вписавши одне слово:

Імовірність події А, обчислену в припущенні, що подія В вже відбулася, називають _______________________________________ імовірністю події А.

1

Завдання № 20

Обведіть кружком правильну відповідь.

Сума двох несумісних подій визначається за формулою:

1) $P(A+B)=P(A)+P(B)$; 2) $P(A+B)=P(A)+P(B)-P(A)\cdot P(B)$

5

Завдання № 21

Розв'яжіть задачу:

Консультаційний пункт інституту отримує пакети з контрольними роботами з міст А, В і С. Імовірність отримання пакета з міста А дорівнює 0,7, з міста В – 0,2. Знайти імовірність того, що черговий пакет буде отриманий з міста С.

$P(C)=$_______________.

1

Завдання № 22

Вставте в дужки відповідну цифру.

Якщо події A_1, A_2, … , A_i, …, A_n утворюють повну групу несумісних подій, то сума їхніх імовірностей дорівнює $\sum_{i=1}^{n}P(A_i)=[\ ]$.

2

Завдання № 23

Обведіть кружком правильну відповідь і дайте пояснення умовного позначення.

1) Імовірність добутку двох залежних подій визначається за формулою:

А) $P(AB)=P(A)*P(B|A)$; Б) $P(AB)=P(A)*P(B)$

2) $P(B|A)$ називається ____

5

Завдання № 24

Розвяжіть задачу:

Із шістьох карток складене слово «ракета». Картки перемішують, потім беруть по одній навмання. Знайти імовірність того, що в результаті випадкового вибору карток знову складеться слово «ракета».

$P(A)=$______________________.

2

Завдання № 25

Напишіть формулу для розв'язання наступної задачі

Імовірність появи події А залежить від подій H_1, H_2, H_3. Визначте повну імовірність події А

$P(A)=$____.

3

Завдання № 26

Розв'яжіть задачу:

Є дві однакові на вигляд урни. У першої знаходяться 2 білі й 3 чорних кулі, у другій - 4 білі і 1 чорна. Навмання з однієї з урн виймають одну кулю. Визначити імовірність того, що вийнята куля виявиться білою.

$P(A)=$____.

4

Завдання № 27

Напишіть формулу для розв'язання наступної задачі

Апріорні імовірності подій H_1, H_2, H_3 відомі. У результаті досліду з'явилася подія А.

Визначте апостеріорну імовірність події H_2.

$P(H_2\,|\,A)=$____.

2

Завдання № 28

Назвіть методи (формули) розв'язання наступної задачі.

1. У кожному з n незалежних випробувань подія А відбувається з постійною імовірністю p. Знайти імовірність того, що подія А відбудеться рівно k разів.

____

2. Як називається ця формула?

____

1

Завдання № 29

Доповніть твердження, вписавши одне слово:

В n незалежних дослідах число появ події А, для якого імовірність перевищує або, принаймні, не менше імовірності кожного з інших можливих чисел появи події А, називається ______________________ числом появи події А.

1

Завдання № 30

Запишіть словосполучення:

Що визначається виразом

$$np - q \le m_0 \le np + p?$$

m_0 - це______________________________________

4

Завдання № 31

2) Визначте відповідність у вигляді комбінації цифр і літер

1) Якщо np - ціле число, то; 2) Якщо np - q - ціле число, то; 3) Якщо np-q - дробове число, то.	А) m_0=np; Б) існують два значення m_0 = np - q і m_0'= np+p; В) існує одне ціле, укладене між дробовими числами

1. _______________ 2.______________ 3.____________

2

Завдання № 32

Обґрунтуйте відповідь одним реченням:

При дуже великому значенні числа дослідів n застосування формули Бернуллі стає неефективним, і замість неї використовують асимптотичну формулу, яку дає локальна теорема Лапласа, тому що в цьому випадку застосування формули Бернуллі приводить до

.____

4

Завдання № 33

Запишіть формулу, продовжіть твердження і визначіть відповідність у вигляді комбінації цифр і літер:

Асимптотична формула локальної теореми Лапласа має вигляд:

____

і дозволяє визначити імовірність того, що

____

де

1) x А) $\dfrac{m - np}{\sqrt{npq}}$

2) $\varphi(x)$ Б) $\dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

1._______________ 2.________________________

7

Завдання № 34

Запишіть формулу, продовжіть твердження і визначіть відповідність у вигляді комбінації цифр і літер:

Формула інтегральної теореми Лапласа має вигляд:

.____

і дозволяє визначити імовірність того, що

____

де

1) x' А) число дослідів

2) x'' Б) імовірність непояви події в одному досліді

3) n В) $\dfrac{m_1 - np}{\sqrt{npq}}$

4) p Г) імовірність появи події в одному досліді

5) q Д) $\dfrac{m_2 - np}{\sqrt{npq}}$.

1._________ 2._________ 3.___________ 4.___________ 5.____________

5

Завдання № 35

Запишіть формулу й перелічіть властивості функції:

Формула інтеграла імовірностей має вигляд:

.____

і має такі властивості:

1) Ф(-x) =____

2) Ф(0) =______________________________

3) Ф(∞) =______________________________

4) Ф(5) =______________________________

4

Завдання № 36

Розв'яжіть задачу:

Імовірності влучення в ціль при стрілянині з першого й другого знарядь відповідно дорівнюють: p_1=0,7; p_2=0,8. Знайти імовірність влучення при одному залпі (з обох знарядь) хоча б одним із знарядь (розв'язати задачу двома способами).

Перший спосіб: ______________________________

Другий спосіб: ______________________________

1

Завдання № 37

Доповніть твердження, вписавши одне слово:

Випадкова величина, число значень якої скінченне або нескінченне, але рахункове (яка може приймати тільки окремі значення) називається

____

1

Завдання № 38

Доповніть твердження, вписавши одне слово:

Випадкова величина, число значень якої нескінченне навіть на невеликому інтервалі називається ______________________________

1

Завдання № 39

Доповніть твердження, вписавши одне слово:

Будь-яке правило, що дозволяє знаходити імовірності різних подій, пов'язаних з випадковою величиною, називається ______________________________

2

Завдання № 40

Доповніть ствердження, вписавши словосполучення і цифру:

Таблиця, у верхньому рядку якої перелічені всі значення випадкової величини $x_1, x_2, \ldots, x_n$ у порядку їхнього зростання, а в нижньому - імовірності появи цих значень $p_1, p_2, \ldots, p_n$: називається ______________________________, причому

$\sum p_i =$ __________

5

Завдання № 41

Визначте відповідність у вигляді комбінації цифр і літер.

1) Випадкова величина;	А) $F(x)$
2) Поточне значення випадкової величини;	Б) $P\{x_1 \le X < x_2\}$
3) Функція розподілу;	В) X
4) Щільність розподілу;	Г) x
5) Імовірність влучення на інтервал;	Д) $f(x)$

1.__________ 2.__________ 3.__________ 4.__________ 5.__________

2

Завдання № 42

Доповніть твердження, вписавши словосполучення й формулу:

Імовірність того, що випадкова величина X прийме значення менше деякого поточного, називається ______________________________

____

5

Завдання № 43

Перелічіть властивості функції розподілу:

1) Значення функції розподілу належать відрізкові ______________

2) Якщо $x_2 > x_1$

____

3) $F(-\infty) =$ ______________________________

4) $F(+\infty) =$

____

5) $P\{x_1 \le X \le x_2\} =$

____

2

Завдання № 44

Доповніть твердження, вписавши словосполучення і формулу:

Щільністю розподілу випадкової величини X у точці x називається ______

____

____

4

Завдання № 45

Перелічіть властивості щільності розподілу:

1) Щільність розподілу є невідємною ____________________________

2) $$\int_{-\infty}^{+\infty} f(x)dx = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_$$

3) $$\int_{-\infty}^{x} f(x)dx = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_$$

4) $$P\{x_1 \le X \le x_2\} = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_$$

5

Завдання № 46

Визначте відповідність у вигляді комбінації цифр і літер.

1) Функція розподілу; А) $F(x)$

2) Щільність розподілу; Б) диференціальний закон

В) інтегральний закон

Г) $f(x)$

1.______________________ 2.________________________

1

Завдання № 47

Обведіть кружком правильну відповідь:

Чи може функція розподілу при якому-небудь значенні аргументу бути відємною?

Так Ні

1

Завдання № 48

Обведіть кружком правильну відповідь:

Чи може функція розподілу при якому-небудь значенні аргументу бути більше одиниці?

Так Ні

1

Завдання № 49

Обведіть кружком правильну відповідь:

Чи може щільність розподілу при якому-небудь значенні аргументу бути від'ємною?

Так Ні

1

Завдання № 50

Обведіть кружком правильну відповідь:

Чи може щільність розподілу при якому-небудь значенні аргументу бути більше одиниці?

Так Ні

5

Завдання № 51

Визначте відповідність у вигляді комбінації цифр і літер.

Яку розмірність має:

1) Функція розподілу;	А) Розмірність квадрата випадкової величини;
2) Щільність розподілу;	Б) Розмірність випадкової величини;
3) Математичне сподівання;	В) Зворотна розмірності випадкової величини;
4) Дисперсія;	
5) Середнє квадратичне відхилення;	Г) Безрозмірна.

1._________ 2._________ 3.___________ 4.___________ 5.____________

2

Завдання № 52

Доповніть твердження, вписавши словосполучення і формулу:

Сума добутків всіх можливих значень випадкової величини Х на імовірності цих значень називається

____

3

Завдання № 53

Визначте відповідність у вигляді комбінації цифр і літер.

1)Математичне сподівання; А) найбільш імовірне значення випадкової величини

2)Мода;	Б) середнє значення випадкової величини
3)Медіана;	В) значення x_m, при якому $P\{X<x_m\}=P\{X>x_m\}=0,5$

1.______________________ 2._________________ 3._________________

10

Завдання № 54

Визначте відповідність у вигляді комбінації цифр і літер.

1) Початковий момент дискретної випадкової величини;
2) Початковий момент безперервної випадкової величини;
3) Початковий момент s-го порядку;
4) Центральний момент s-го порядку;
5) Центрування випадкової величини;
6) Центральний момент дискретної випадкової величини;
7) Центральний момент безперервної випадкової величини;
8)Дисперсія;
9)Математичне сподівання.

А) $\overset{0}{X} = X - m_x$

Б) $\alpha_s = \sum_{i=1}^{n} x_i^s p_i$

В) $\alpha_s = \int_{-\infty}^{+\infty} x^s f(x)dx$

Г) $\alpha_s = M[X^s]$

Д) $\mu_s = M[\overset{0}{X}{}^s]$

Е) $\mu_s = \sum_{i=1}^{n} (x_i - m_x)^s p_i$

Ж) $\mu_s = \int_{-\infty}^{+\infty} \overset{0}{x}{}^s f(x)dx$

З) $\int_{-\infty}^{+\infty} \overset{0}{x}{}^2 f(x)dx$

І) $\int_{-\infty}^{+\infty} xf(x)dx$

1._____ 2._____ 3._____ 4._____ 5._____ 6.____ 7._____ 8._____ 9._____

5

Завдання № 55

Назвіть властивості числових характеристик, дописавши вирази:

1) Математичне сподівання невипадкової величини с $M[c] =$ _______________
2) Математичне сподівання добутку невипадкової величини с на випадкову величину Х $M[cX] =$ ____
3) Дисперсія невипадкової величини с $D[c] =$ _____________________________
4) Дисперсія добутку невипадкової величини с на випадкову величину Х $D[cX] =$ ____

1

Завдання № 56

Доповніть твердження, вписавши одне слово:

Виразом $M[X^2] - (M[X])^2$ визначається ____________________________

6

Завдання № 57

Визначте відповідність у вигляді комбінації цифр і літер.

1) розподіл Бернуллі;
2) закон Пуассона;
3) експонентний закон;
4) нормальний закон;
5) рівномірний розподіл.

А) $F(x) = \frac{x-\alpha}{b-a}$

Б) $f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left\{\frac{(x-m)^2}{2\sigma^2}\right\}$

В) $f(t) = \lambda e^{-\lambda t}$

Г) $P(k) = \frac{(\lambda\tau)^k}{k!} e^{-\lambda\tau}$

Д) $P_n(m) = C_n^m p^m q^{n-m}$

1. _____ 2._____ 3._____ 4._____ 5._____

4

Завдання № 58

Визначите відповідність математичних сподівань законам розподілу у вигляді комбінації цифр і букв.

1) розподіл Бернуллі;
2) Закон Пуассона;
3) експонентний закон;
4) рівномірний розподіл;

А) $\lambda\tau$;
Б) $\dfrac{b+a}{2}$;
В) np;
Г) $1/\lambda$;

1. _____ 2._____ 3._____ 4._____

4

Завдання № 59

Визначте відповідність дисперсій законам розподілу у вигляді комбінації цифр і літер.

1) розподіл Бернуллі;
2) закон Пуассона;
3) експонентний закон;
4) рівномірний розподіл;

А) $1/\lambda^2$,
Б) npq;
В) $\lambda\tau$;
Г) $\dfrac{(b-a)^2}{12}$

1._____ 2._____ 3._____ 4._____

8

Завдання № 60

Розв'яжіть задачу:

Двічі кидають монету. Випадкова величина Х - число появи гербів.
1) Побудуйте ряд розподілу випадкової величини Х:

x_i			
p_i			

2) Визначте числові характеристики випадкової величини Х (оцініть середнє число появ герба в даному досліді й відхилення від середнього)

$m_x =$ ________________________ $D_x =$ ______________________________ $\sigma_x =$
.
3) При двох киданнях монети число появ герба в середньому дорівнює __________

3

Завдання № 61

Доповніть твердження, вписавши одне слово й формулу:

Якщо число дослідів n нескінченно велике, а імовірність появи події в одному досліді р дуже мала, біноміальний розподіл заміняють законом _____________________________________, що визначається виразом

____

8

Завдання № 62

Визначте закони розподілу випадкової величини Х - число викликів і випадкової величини Т - проміжок часу між викликами, розвяжіть задачу:

На АТС надходять виклики з інтенсивністю λ, яка дорівнює одному виклику у секунду.

1) Визначити імовірність того, що протягом 5 секунд а) не надійде жодного виклику; б) надійде хоча б один виклик.

Число викликів АТС розподілено за законом ________________________________.

а) $P(A) =$________________________ б) $P(B) =$

2) Визначити середнє значення проміжку часу між викликами й відхилення від середнього.

Проміжок часу між викликами Т розподілений за _______________ законом.

Середнє значення проміжку часу між викликами $m_x =$________й відхилення від середнього $\sigma_x =$________.

Обґрунтувати відповідь у словесній формі

____.

5

Завдання № 63

Визначте відповідність у вигляді комбінації цифр і літер для характеристик потоку подій

1) число подій, що потрапляють на інтервал часу;	А) безперервна випадкова величина;	г) τ;
2) проміжок часу між двома випадковими подіями в потоці;	Б) дискретна випадкова величина;	д) k;
3) інтенсивність потоку подій.	В) невипадкова величина.	е) λ.

1.________________ 2.________________ 3.________________

1

Завдання № 64

Доповніть твердження, вписавши одне слово:

Імовірність безвідмовної роботи об'єкта протягом певного часу називається

____

2

Завдання № 65

Доповніть твердження, вписавши формулу:

Формула ________________________________ дозволяє визначити імовірність безвідмовної роботи об'єкта протягом часу t.

2

Завдання № 66

Доповніть твердження, вписавши декілька слів:

Нормальний розподіл визначається двома параметрами - _________________ ______________________________і ____

3

Завдання № 67

Напишіть формулу для розв'язання наступної задачі:

Визначте імовірність влучення нормальної випадкової величини на ділянку значень (x_1,x_2)

____

2

Завдання № 68

Доповніть твердження, вписавши словосполучення:

Вираз $\Phi(x)=\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_0^x e^{-\frac{t^2}{2}}dt$ називається ________________________________

4

Завдання № 69

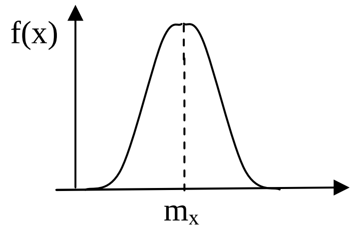
Назвіть чотири властивості інтеграла імовірностей:

1) при x=0 Φ(x)=____;
2) при x=∞ Φ(x)=____;
3) при x=-∞ Φ(x)=____;
4) Φ(-x)=____.

4

Завдання № 70

Назвіть найменування закону розподілу, запишіть його аналітичне вираження:



найменування:

____

формула:

____

5

Завдання № 71

Напишіть формулу для розв'язання наступної задачі

Випадкова величина X розподілена нормально з параметрами m_x=1, σ_x=0,2. Визначити імовірність влучення X у інтервал (0,8;1,4)

$P\{0{,}5 < X < 1{,}5\} =$ ____.

7

Завдання № 72

Напишіть формулу для розв'язання наступної задачі

Випадкова величина X розподілена нормально з параметрами m_x=8, σ_x=0,5. Визначити імовірність того, що X відхилиться від свого математичного сподівання не більше ніж на 1

$P\left\{\left|\overset{o}{x}\right|<1\right\}=$ ____.

4

Завдання № 73

Доповніть твердження, вписавши числа та назву правила:

1) 68 % значень будь-якої нормально розподіленої випадкової величини належать інтервалу ________________________
2) 95 % значень будь-якої нормально розподіленої випадкової величини належать інтервалу ________________________
3) 99,7 % значень будь-якої нормально розподіленої випадкової величини належать інтервалу ________________________
4) дане правило називається ____

1

Завдання № 74

Як називається теорема (впишіть словосполучення):

Якщо випадкова величина $Y = X_1 + X_2 +\ldots+X_n$, де вплив кожного з доданків на всю

суму рівномірно малий, то величина Y має розподіл, близький до нормального, і тим ближче, чим більше n .
Дана теорема називається____

1

Завдання № 75

Доповніть твердження, вписавши формулу:

Функція розподілу системи двох випадкових величин (X,Y) дорівнює імовірності того, що випадкова величина X прийме значення, менше x і випадкова величина Y прийме значення, менше y:
F(x,y) =____.

3

Завдання № 76

Назвіть властивості щільності розподілу системи (X,Y):

1) $f(x,y)$ ____;

2) $\int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f(x,y)dxdy =$ ____;

3) $\iint\limits_{-\infty}^{+\infty} f(x,y)dxdy =$ ____.

8

Завдання № 77

Назвіть властивості функції розподілу системи (X,Y):

1) Якщо $x_2 > x_1$, $F(x_2,y)$ _____ $F(x_1,y)$;
2) Якщо $y_2 > y_1$, $F(x,y_2)$ _____ $F(x,y_1)$;
3) F(x, - ∞) = __________________________;
4) F(-∞, y) = __________________________;
5) F(-∞, - ∞) = _______________________;
6) F(x, + ∞) = _________________________;
7) F(+∞, y) = __________________________;
8) F(+∞, + ∞) = _______________________.

5

Завдання № 78

Визначте відповідність у вигляді комбінації цифр і літер.

1) Початковий момент системи двох випадкових величин (X, Y);
2) Центральний момент системи випадкових величин (X, Y);
3) Початкові моменти першого порядку;
4) Центральні моменти другого порядку;
5) Другий змішаний центральний момент;
6) Порядок моменту.

А) $\alpha_{1,0} = M[X]\ \alpha_{0,1} = M[Y]$
Б) $\mu_{x,y} = K_{xy}$
В) $\alpha_{k,s} = M[X^k Y^s]$
Г) $\mu_{2,0} = D_x,\ \mu_{0,2} = D_y$
Д) $\mu_{k,s} = M[\overset{0}{X}{}^k \overset{0}{Y}{}^s]$
Е) k+s.

1. _____ 2._____ 3._____ 4._____ 5._____ 6.____

4

Завдання № 79

Доповніть твердження, вписавши словосполучення, допишіть формулу, поясніть її компоненти:

1) Характеристикою зв'язку між величинами X і Y є ______________________
____
2) Допишіть формулу

$$K_{x,y} = \sum_{i=1}^{n}\sum_{j=1}^{m} \overset{0}{x}_i \overset{0}{y}_j p_{ij} = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_$$

____
3) Поясніть p_{ij} -____
____

6

Завдання № 80

Показати, що у випадку незалежних X і Y кореляційний момент дорівнює нулю, перетворивши вираз і дописавши обґрунтування перетворень:

Перетворіть вираз

$$K_{x,y} = \sum_{i=1}^{n}\sum_{j=1}^{m} \overset{0}{x}_i \overset{0}{y}_j p_{ij} = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_$$

Допишіть обґрунтування
____

	4

Завдання № 81

Назвіть характеристику, поясніть компоненти формули:

$$r_{xy}=\frac{K_{xy}}{\sigma_x\sigma_y}$$ - ________________________________

Поясніть компоненти: $K_{x,y}$ - ________________________________

σ_x,σ_y - ________________________________

3

Завдання № 82

Доповніть твердження, вставивши в скобки відповідні цифри й дописавши словосполучення

Коефіцієнт кореляції лежить у межах $[\ ]\le r_{xy}\le[\ ]$ і дорівнює нулю,

якщо ________________________________

5

Завдання № 83

Визначте відповідність у вигляді комбінації цифр і літер.

1) Математичне сподівання суми залежних або незалежних випадкових величин;

2) Математичне сподівання лінійної функції декількох випадкових величин;

3) Математичне сподівання добутку двох випадкових величин;

4) Математичне сподівання добутку некорельованих випадкових величин.

А) $\sum_{i=1}^{n} a_i M[X_i]+b$;

Б) M[X] + M[Y];

В) $\prod_{i=1}^{n} M[X_i]$;

Г) M[X] * M[Y] + K_{xy}.

1. ________ 2.________ 3.________ 4.________

4

Завдання № 84

Визначте відповідність у вигляді комбінації цифр і літер.

1) Дисперсія суми декількох випадкових величин;

2) Дисперсія суми некорельованих випадкових величин;

3) Дисперсія лінійної функції декількох випадкових величин $\sum_{i=1}^{n} a_i X_i + b$;

4) Дисперсія добутку двох незалежних випадкових величин X і Y;

А) $\sum_{i=1}^{n} a_i^2 D[X_i]+2\sum_{i<j} a_i a_j K_{x_i x_j}$;

Б) $\sum_{i=1}^{n} D[X_i]$;

В) $\sum_{i=1}^{n} D[X_i]+2\sum_{i<j} K_{x_i x_j}$;

Г) D[X]*D[Y] + $M[X]^2$D[Y] + $M[Y]^2$D[X].

1. ________ 2.________ 3.________ 4.________

4

Завдання № 85

Розв'яжіть задачу:

Визначити математичне сподівання і дисперсію випадкової величини Z=2+3(X-Y).

Числові характеристики: $m_x=-2$; $m_y=4$; $D_x=4$; $D_y=9$; $r_{xy}=-0{,}5$.

$M(Z)=$ ________________________________.

$D(Z)=$ ________________________________.

3

Завдання № 86

Доповніть твердження, дописавши словосполучення

Імовірність, якою можна зневажити в даному дослідженні, називається

____

і позначається ________________________________

2

Завдання № 87

Як називається теорема, запишіть формулу:

При необмеженому збільшенні числа n незалежних випробувань середня арифметична спостережуваних значень випадкової величини збігається за імовірністю до її математичного сподівання. - ________________________________

Записати формулу________________________________

ТЕСТИ до ЗМ2. ЕЛЕМЕНТИ МАТЕМАТИЧНОЇ СТАТИСТИКИ

1

Завдання № 88

Доповніть твердження, дописавши одне слово:

Числова характеристика, отримана в результаті обробки випадкової вибірки, називається _________________________цієї числової характеристики.

2

Завдання № 89

Доповніть твердження, дописавши позначення

Статистичними аналогами параметрів розподілу є генеральна середня $\overline{X}$ і генеральна дисперсія $\sigma^2_{ген}$. Розраховані за даними вибірки числові характеристики позначають відповідно ____

5

Завдання № 90

Доповніть твердження, дописавши словосполучення

1) Помилки, що виникають через неточності й похибки при одержанні відомостей про одиниці сукупності, називаються ____________________________

2) Помилки, що уявляють собою різницю між вибірковими й генеральними характеристиками досліджуваної сукупності називаються ______ ____

3) Помилки репрезентативності підрозділяють на _____________________________

4) Помилки, пов'язані з тим, що структура вибірки відрізняється від структури генеральної сукупності, називаються_____________________________________

5) Помилки, які пояснюються тим, що досліджується тільки частина сукупності, називаються ____

1

Завдання № 91

Доповніть твердження, дописавши словосполучення

Сукупність спостережуваних предметів або явищ, об'єднаних якою-небудь загальною ознакою або властивістю якісного або кількісного характеру називається ____

6

Завдання № 92

Визначте відповідність у вигляді комбінації цифр і літер.

1) Число одиниць вибіркової або генеральної сукупності; 2) Кожне окреме значення досліджуваної ознаки; 3) Число, що показує, скільки разів зустрічається в сукупності та або інша варіанта.	А) Обсяг сукупності; Б) Варіанта; В) Частота.	а) $x_1, x_2, \ldots, x_n$; б) $m_1, m_2, \ldots, m_n$; в) n.

1. ________ 2._______ 3._______

1

Завдання № 93

Доповніть твердження, вписавши словосполучення

Таблиця, в одному рядку якої розташовуються варіанти $x_1, x_2, \ldots, x_n$ у зростаючому або убутному порядку, а в іншій — відповідні ним частоти $m_1, m_2, \ldots, m_n$, називається ____

2

Завдання № 94

Доповніть твердження, вписавши словосполучення

1) Варіація, при якій окремі значення ознаки (варіанти) відрізняються одне від одного на деяку скінченну величину (звичайно ціле число), називається ____
2) Варіація, при якій значення ознаки можуть відрізнятися одне від іншого на як завгодно малу величину, називається ______________________________

4

Завдання № 95

Визначте відповідність у вигляді комбінації цифр і літер:

1) Величина ознаки відкладається на осі   А) Полігон розподілу;

абсцис, частоти - на осі ординат;

2) На осі абсцис відкладають значення ознаки, а ординатами слугують вертикальні відрізки, які дорівнюють накопиченій відносній частоті варіант;	Б) Гістограма;
3) На осі абсцис відкладають відрізки, що відповідають інтервалам, на яких будують прямокутники із площею, пропорційною відносним частотам інтервалів.	В) Кумулятивна крива.

1. _________ 2.________ 3.________

5

Завдання № 96

Установіть правильну послідовність дій при визначенні закону розподілу за статистичним даними, показавши порядок цифрами:

☐ Становлять варіаційний ряд у графічному вигляді (будують кумулятивну криву або гістограму);

☐ Визначають вид закону розподілу з теоретичних міркувань або із графіка;

☐ Визначають відносні частоти влучення досліджуваної ознаки в інтервал;

☐ Перевіряють гіпотезу про передбачуваний закон розподілу досліджуваної ознаки;

☐ Визначають вибіркові параметри розподілу за методом моментів або найменших квадратів;

Представляють статистичні дані у вигляді групованого варіаційного ряду.

4

Завдання № 97

Доповніть твердження, вписавши словосполучення

1) Різниця між найбільшим і найменшим значеннями варіаційного ряду $R = x_{max} – x_{min}$ називається____

2) Середнє арифметичне абсолютних значень відхилень варіант від середньої d називається ____

3) Середнє арифметичне квадрата відхилень значень ознак ряду від їх середнього арифметичного σ^2 називається ______________________________

4) Доля стандартного відхилення від середнього значення варіаційного ряду V називається ____

3

Завдання № 98

Доповніть твердження, вписавши кілька слів:

Оцінки числових характеристик повинні мати властивості ____

1

Завдання № 99

Вибрати правильну відповідь, обвівши кружком цифру

Незміщеність оцінки параметра a* означає:

1) що вона характеризується найменшою дисперсією;

2) що її математичне сподівання дорівнює математичному сподіванню випадкової величини Х;

3) збільшення її точності із збільшенням обсягу вибірки.

1

Завдання № 100

Вибрати правильну відповідь, обвівши кружком цифру

Ефективність оцінки параметра a* означає:

1) що вона характеризується найменшою дисперсією;

2) що її математичне сподівання дорівнює математичному сподіванню випадкової величини Х;

3) збільшення її точності із збільшенням обсягу вибірки.

1

Завдання № 101

Вибрати правильну відповідь, обвівши кружком цифру

Спроможність оцінки параметра a* означає:

1) що вона характеризується найменшою дисперсією;

2) що її математичне сподівання дорівнює математичному сподіванню випадкової величини Х;

3) збільшення її точності із збільшенням обсягу вибірки.

3

Завдання № 102

Визначте відповідність у вигляді комбінації цифр і літер:

1) Якщо оцінка параметра a^* при $n \to \infty$ збігається за імовірностю до оцінюваного параметра a, то вона є; — А) спроможною;

2) Якщо математичне сподівання оцінки параметра a^*дорівнює оцінюваному параметрові a $M[a^*] = a$, то вона є; — Б) ефективною;

3) Якщо при заданому обсязі вибірки оцінка параметра a^* має найменшу дисперсію, то вона є. — В) незміщеною.

1. ________ 2._______ 3._______

2

Завдання № 103

Доповніть твердження, вписавши кілька слів:

Для визначення помилки від заміни генерального параметра його оцінкою використовують

____

4

Завдання № 104

Установіть відповідність у вигляді комбінації цифр і літер:

1) Залежна змінна;	А) Результативна ознака;	д) Y;
2) Незалежна змінна.	Б) Пояснювальна змінна;	
	В) Пояснююча змінна;	е) X.
	Г) Ознака-фактор.	

1. ___________________ 2.___________________

3

Завдання № 105

Визначте відповідність у вигляді комбінації цифр і літер:

1) Залежність $y = f(x)$, у якій кожному значенню X відповідає одне певне значення Y; — А) Функціональна;

2) Залежність, в якій одному значенню змінної X x_i відповідає ряд значень змінної Y: $y_{i1}, y_{i2}, \ldots, y_{ik}$; — Б) Регресія;

3) Умовне математичне сподівання випадкової величини Y за умови, що X прийняла значення x_i; — В) Кореляційна;

4) Залежність, в якій при зміні X змінюється середнє значення Y. — Г) Статистична.

1. ________ 2._______ 3._________ 4.________

2

Завдання № 106

Доповніть твердження, вписавши кілька слів і формулу:

При використанні методу найменших квадратів вимога найкращого узгодження $\bar{y}_x = \varphi(x)$ з дослідними даними зводиться до того, щоб

____,

що виражається формулою:

____

1

Завдання № 107

Вибрати правильну відповідь, обвівши кружком цифру

Суть методу найменших квадратів перебуває в:

1) мінімізації суми залишкових величин;

2) мінімізації дисперсії результативної ознаки;

3) мінімізації суми квадратів залишкових величин.

6

Завдання № 108

Установіть правильну послідовність розрахунків, застосовуваних при використанні методу найменших квадратів:

☐ Суму квадратів відхилень кривої, що згладжує залежність, від експериментальних точок записують як функцію аргументу x і невідомих параметрів a_j;

☐ Визначають клас залежності $\bar{y}_x = \varphi(x)$;

☐ В результаті перетворення системи отримують значення параметрів a_j;

☐ Беруть похідні від $\sum_{i=1}^{n}[\varphi(x_i, a_1, a_2, \ldots, a_j, \ldots, a_m) - y_i]^2 \to \min$ за параметрами a_j і дорівнюють їх нулю, в результаті чого отримують систему рівнянь;

☐ Перевіряють статистичну значимість залежності $\bar{y}_x = \varphi(x)$ і параметрів a_j;

☐ Записують залежність $\bar{y}_x = \varphi(x)$ з використанням отриманих параметрів a_j.

6

Завдання № 109

Доповніть твердження, вписавши словосполучення:

1) Вибірковий коефіцієнт кореляції $r_в$ служить для ________________________
____
2) Вибірковий коефіцієнт кореляції $r_в$ приймає значення
____
3) Якщо $r_в = 0$, то лінійний зв'язок
____
4) Якщо $r_в = 1$ то лінійний зв'язок
____
5) Вибіркове кореляційне відношення η застосовують для _________________
____
6) Вибірковим кореляційним відношенням Y до X називається _____________
____

3

Завдання № 110

Визначіть відповідність у вигляді комбінації цифр і літер:

1) Дисперсія групових середніх щодо загальної середньої;	А) Загальна дисперсія;
2) Середнє арифметичне групових дисперсій;	Б) Внутрігрупова дисперсія;
3) Дисперсія ознаки Y;	В) Міжгрупова дисперсія.

1. ________ 2.________ 3.__________

6

Завдання № 111

Впишіть відповідні цифри й символи:

1) Якщо Y функціонально залежить від X, то внутрішньогрупова дисперсія $D_{вн.гр} =$ _____, загальна дисперсія Y D_y =_________, а кореляційне відношення дорівнює $\eta_{yx} =$ ________;

2) Якщо між X і Y зв'язок відсутній, то міжгрупова дисперсія $D_{межгр} =$ _____, загальна дисперсія Y D_y =________, а кореляційне відношення дорівнює $\eta_{yx} =$ ____.

1

Завдання № 112

Доповніть твердження, вписавши в скобки відповідні цифри:

Значення кореляційного відношення лежить у межах $[\ ] \le \eta \le [\ ]$.

2

Завдання № 113

Доповніть твердження, обвівши кружальцем відповідну цифру:

Значення кореляційного відношення η

1) більше або дорівнює

2) менше або дорівнює

вибірковому коефіцієнту кореляції $r_в$

η ____ $|r_в|$

2

Завдання № 114

Доповніть твердження, вписавши одне слово:

Якщо кореляційне відношення дорівнює вибірковому коефіцієнту кореляції, $\eta = |r_в|$, то між X і Y існує __________________________ кореляційна залежність.

1

Завдання № 115

Доповніть ствердження, вписавши в скобки відповідні цифри:

У практичних задачах при перевірці гіпотез найчастіше призначають рівень значущості $\alpha = [\_\_\_\_\_\_\_\_\_\_]$.

1

Завдання № 116

Доповніть твердження, вписавши словосполучення:

Область значень критерію Z, що відповідають рівню значущості α, називають
____.

1

Завдання № 117

Доповніть твердження, вписавши словосполучення:

Область значень критерію Z, що відповідають імовірності $1-\alpha$, називають
____.

1

Завдання № 118

Доповніть твердження, вписавши словосполучення і позначення:

Значення критерію, що відокремлює область прийняття гіпотези від критичної області називається ____

2

Завдання № 119

Установіть відповідність у вигляді комбінації цифр і літер:

1) При перевірці гіпотези зроблено помилку першого роду;
2) При перевірці гіпотези зроблено помилку другого роду.

А) Відкинута нульова гіпотеза, у той час як вона була правильною;
Б) Прийнята нульова гіпотеза, у той час як вона була невірною.

1. ________ 2.________

2

Завдання № 120

Доповніть твердження, вписавши кілька слів і позначення:

Потужністю критерію називається імовірність ______________________________,
яка позначається ______________________________.

2

Завдання № 121

Доповніть твердження, дописавши вираження:

При визначенні двосторонньої критичної області найбільша потужність критерію досягається, якщо імовірність влучення критерію Z у кожен із двох інтервалів критичної області дорівнює $P\{|Z|>z_{кр}\}$=____________________.

2

Завдання № 122

Доповніть ствердження, дописавши в скобки інтервал:

При перевірці гіпотези про значущість розбіжності між вибірковою середньою $\tilde{X}$ й генеральною середньою $\overline{X}$ нормальної сукупності при генеральній дисперсії $\sigma^2_{ген}$ розподіл критерію Z симетричний відносно нуля, тому область прийняття нульової гіпотези (______________________________).

6

Завдання № 123

Установіть відповідність у вигляді комбінації цифр і літер:

При оцінці генеральній середньої $\overline{X}$ по вибірковій середній $\tilde{X}$ визначити обсяг вибірки, який забезпечить задану припустиму величину помилки δ можна за формулою $n = \dfrac{t^2_{двуст}(\alpha,k)*\sigma^2_{выб}}{(\overline{X}-\tilde{X})^2}$, де

1) $t^2_{двуст}$;
2) $(\overline{X}-\tilde{X})$;
3) α;
4) k;
5) $\sigma^2_{выб}$.

А) Рівень значущості;
Б) Вибіркове середнє квадратичне відхилення;
В) Число ступенів свободи;
Г) Величина помилки δ;
Д) Значення критерію.

1. ________ 2.________ 3.________ 4.________ 5.__________

2

Завдання № 124

Доповніть твердження, дописавши формулювання гіпотез:

Для перевірки значущості розбіжності між вибірковою середньою $\overline{X}$ і генеральної середньою $\tilde{X}$ нормальної сукупності формулюють гіпотези:
нульову ______________________________
і альтернативну ______________________________.

3

Завдання № 125

Доповніть твердження, дописавши нульову гіпотезу, кілька слів або формулу:

У якості критерію для перевірки нульової гіпотези ________________________ про рівність дисперсій нормальних генеральних сукупностей приймають випадкову величину, яка являє собою відношення:
______________________________.

2

Завдання № 126

Доповніть твердження, дописавши кілька слів і позначення:

Розподіл Фішера, який має випадкова величина F, залежить тільки від ______________________________.

2

Завдання № 127

Доповніть твердження, дописавши вираз:

При порівнянні вибіркової $\sigma^2_{виб}$ й генеральної дисперсій нормальної сукупності перевіряють гіпотезу про те, що генеральна дисперсія дорівнює гіпотетичному

значенню σ^2_0. Нульову гіпотезу в цьому випадку записують в такий спосіб:
____.

2

Завдання № 128

Доповніть твердження, дописавши позначення:

При порівнянні вибіркової $\sigma^2_{виб}$ і генеральної дисперсій нормальної сукупності для перевірки нульової гіпотези про те, що генеральна дисперсія дорівнює гіпотетичному значенню σ^2_0, використовують критерій________________
з ________________ ступенями свободи.

2

Завдання № 129

Доповніть твердження, дописавши позначення або назву:

Для перевірки гіпотези про нормальний розподіл генеральної сукупності застосовується критерій ____

7

Завдання № 130

Установіть правильну послідовність розрахунків, застосовуваних при перевірці гіпотези за допомогою критерію χ^2, показавши порядок цифрами:

- ☐ Визначають $\chi^2_{критич}$;
- ☐ Знаходять емпіричні частоти влучення значень досліджуваної ознаки в інтервал;
- ☐ Обчислюють значення χ^2;
- ☐ Знаходять теоретичні частоти влучення значень досліджуваної ознаки в інтервали;
- ☐ Знаходять теоретичні значення імовірностей влучення досліджуваної ознаки в інтервали;
- ☐ Порівнюють значення χ^2 і $\chi^2_{критич}$;
- ☐ Розраховують оцінки числових характеристик.

1

Завдання № 131

Доповніть твердження, дописавши словосполучення або одне слово:

Дисперсійний аналіз як метод математичної статистики застосовують для оцінки впливу __________________________ на результати спостережень.

2

Завдання № 132

Доповніть твердження, дописавши одне слово або словосполучення й формулу:

Розсіювання за фактором (розсіювання між групами) оцінюється за допомогою суми квадратів відхилень групових середніх від загальної середньої, котра називається ____
і визначається за формулою ____.

2

Завдання № 133

Доповніть твердження, дописавши одне слово або словосполучення і формулу:

Сума квадратів відхилень спостережуваних значень ознаки X від своєї групової середньої $\overline{x}_{грj}$ називається ____
і визначається за формулою: ____.

2

Завдання № 134

Доповніть твердження, дописавши формулу:

Загальна сума квадратів відхилень спостережуваних значень ознаки X від своєї середньої $\overline{X}$ визначається за формулою: $S_{заг}$ = _________________________

5

Завдання № 135

Доповніть твердження, дописавши формули:

Використовуючи загальну, факторну і залишкову суми квадратів відхилень, визначають відповідні дисперсії, причому число ступенів свободи
загальної дисперсії дорівнює ______________________________________
факторної дисперсії дорівнює ______________________________________

залишкової дисперсії дорівнює ____

1

Завдання № 136

Вибрати правильну відповідь, обвівши кружком цифру

Факторна сума квадратів відхилень в лінійній парній моделі має число ступенів свободи, яке дорівнює:

1) $n-1$;

2) 1;

3) $n-2$.

1

Завдання № 137

Вибрати правильну відповідь, обвівши кружком цифру

Залишкова сума квадратів відхилень у лінійній парній моделі має число ступенів свободи, яке дорівнює:

1) $n-1$;

2) 1;

3) $n-2$.

1

Завдання № 138

Вибрати правильну відповідь, обвівши кружком цифру

Загальна сума квадратів відхилень у лінійній парній моделі має число ступенів свободи, що дорівнює:

1) $n-1$;

2) 1;

3) $n-2$.

1

Завдання № 139

Вибрати правильну відповідь, обвівши кружком цифру

Для оцінки значущості коефіцієнтів регресії розраховують:

1) F-критерій Фішера;

2) t-критерій Стьюдента;

3) коефіцієнт детермінації r_{xy}^2.

2

Завдання № 140

Доповніть твердження, вписавши слово або позначення і формулу:

Для оцінки впливу фактора формулюють нульову гіпотезу про рівність факторної і залишкової дисперсій, перевірку якої здійснюють за критерієм ____________________, що визначається за формулою __________________.

5

Завдання № 141

Установіть відповідність у вигляді комбінації цифр і літер:

1) Нульова гіпотеза про рівність факторної і залишкової дисперсій помилкова; 2) Нульова гіпотеза про рівність факторної і залишкової дисперсій справедлива.	А) Факторна дисперсія менше залишкової; Б) Факторна дисперсія більше залишкової.	а) $F_{набл} > F_{кр}$; б) $F_{набл} < F_{кр}$.

1. ____________________ 2.________________________

1

Завдання № 142

Доповніть твердження, вписавши словосполучення:

Оцінка значущості рівняння регресії в цілому і його окремих параметрів є задачею

____

1

Завдання № 143

Доповніть твердження, вписавши словосполучення або позначення:

Перевірка адекватності рівняння регресії експериментальним даним виконується за критерієм ________________________, при цьому висувається нульова гіпотеза ____.

1

Завдання № 144

Вибрати правильну відповідь, обвівши кружком цифру

Значущість рівняння регресії в цілому оцінює

1) F - критерій Фішера;

2) t - критерій Стьюдента;

3) коефіцієнт детермінації r_{xy}^2.

1

Завдання № 145

Вибрати правильну відповідь, обвівши кружком цифру

Залишкова сума квадратів дорівнює нулю

1) коли правильно підібрана регресійна модель;

2) коли між ознаками існує точний функціональний зв'язок;

3) ніколи.

8

Завдання № 146

Установіть відповідність у вигляді комбінації цифр і літер:

Для статистичних розрахунків використовуються функції Microsoft Excel

1) СРЗНАЧ;	А) Оцінює дисперсію за вибіркою;
2) ДИСП;	Б) Дає значення параметра ρ_{yx};
3) СТАНДОТКЛОН;	В) Повертає середнє добутків відхилень для кожної пари точок даних;
4) КОВАР;	Г) Дає значення параметру b;
5) КОРРЕЛ;	Д) Повертає довірчий інтервал для середньої генеральної сукупності;
6) ДОВЕРИТ;	Е) Повертає коефіцієнт кореляції;
7) НАКЛОН;	Ж) Повертає середнє арифметичне своїх аргументів;
8) ОТРЕЗОК.	З) Оцінює середнє квадратичне відхилення.

1. ______ 2.______ 3.______ 4.______ 5._______ 6._______ 7.______ 8.______

2

Завдання № 147

Доповніть твердження, вписавши комбінацію клавіш:

Для визначення параметрів лінійної залежності між X і Y використовується функція Microsoft Excel ЛИНЕЙН, яка повинна задаватися у вигляді формули масиву. Для завершення вводу функції використовується команда масиву: ____

6

Завдання № 148

Доповніть твердження, дописавши цифри і слова:

Для визначення коефіцієнтів ρ_{yx} і b за допомогою функції Microsoft Excel ЛИНЕЙН(известные_значения_y;известные_значения_x;конст;статистика) необхідно виконати наступні дії:

1) Виділити в аркуші Excel ______________________________ комірки;

2) Як аргумент известные _значения_y задають групу комірок ____

3) Як аргумент известные _значения_x задають групу комірок ____

4) Як аргумент конст необхідно задати______________________________

5) Як аргумент статистика необхідно задати ________________________

6) Завершити введення натисканням клавіш ____________________________

1

Завдання № 149

Доповніть твердження, дописавши одне слово:

Для отримання ряду значень y, обчислених на підставі апроксимуючої залежності $y=\rho_{xy}$ + b, використовується функція Microsoft Excel ____________________

1

Завдання № 150

Доповніть твердження, дописавши одне слово:

Для перевірки статистичної значущості рівняння лінійної регресії і його

параметрів використовується функція Microsoft Excel ________________________.
Як аргумент Статистика зазначеної функції вводиться значення _______________.
Введення функції завершується комбінацією клавіш __________________________.

6

Завдання № 151

Установіть відповідність у вигляді комбінації цифр і літер:

З метою проведення регресійного аналізу використовують функцію Microsoft Excel ЛИНЕЙН, в результаті дістають параметри і статистики:

1) ρ_{yx} і b;
2) m_ρ і m_b;
3) m_y;
4) r^2;
5) k;
6) $(y_{ip} - \bar{y})^2$;
7) $(y_i - y_{ip})^2$;

А) значення параметрів лінійної регресії;
Б) стандартні помилки параметрів;
В) стандартна помилка для прогнозних значень у;
Г) коефіцієнт детермінації;
Д) число ступенів свободи;
Е) факторна сума квадратів відхилень;
Ж) залишкова сума квадратів відхилень;

1. ______ 2.______ 3.______ 4.______ 5._______ 6._______

2

Завдання № 152

Доповніть твердження, дописавши по одному слову:

Функція Microsoft Excel НОРМРАСП(х;среднее;стандартное_откл;интегральная) повертає нормальну функцію розподілу для зазначеного середнього і стандартного відхилення, якщо аргумент Интегральная має значення ______________________________ і повертає щільність розподілу, якщо аргумент Интегральная має значення ____________________.

1

Завдання № 153

Доповніть твердження, дописавши одне слово:

Табличне значення F-критерію можна визначити за допомогою статистичної функції Microsoft Excel ____.

1

Завдання № 154

Доповніть твердження, дописавши одне слово:

Табличне значення t-критерію можна визначити за допомогою статистичної функції Microsoft Excel ____.

2

Завдання № 155

Доповніть твердження, дописавши одне слово:

Функція X аргументу t називається випадковою, якщо при кожному заданому значенні аргументу t величина X є

________________________________.

1

Завдання № 156

Доповніть твердження, дописавши одне слово:

Вигляд, прийнятий випадковою функцією X у насчлідку досліду, називається ________________________________ випадкової функції X.

1

Завдання № 157

Доповніть твердження, дописавши два слова:

Імовірнісними характеристики нестаціонарного випадкового процесу є

____.

1

Завдання № 158

Доповніть твердження, дописавши два слова:

Перетином випадкової функції, який відповідає даному t, є

____.

1

Завдання № 159

Доповніть твердження, дописавши одне слово:

Характеристики випадкових функцій являють собою функції

____.

1

Завдання № 160

Виберіть правильну відповідь, обвівши кружком цифру

Математичним сподіванням випадкової функції X(t) називається

1) невипадкова функція;

2) випадкова функція

$m_x(t)$, що при кожному значенні аргументу t дорівнює математичному сподіванню відповідного перетину випадкової функції. $m_x(t) = M[X(t)]$.

1

Завдання № 161

Виберіть правильну відповідь, обвівши кружком цифру

Дисперсією випадкової функції X(t) називається

1) невипадкова функція;

2) випадкова функція

$D_x(t)$, значення якої для кожного t дорівнює дисперсії відповідного перетину випадкової функції $D_x(t) = D[X(t)]$.

1

Завдання № 162

Доповніть твердження, дописавши два слова:

Ступінь залежності між перетинами випадкової функції, що належать до різних t характеризує____.

2

Завдання № 163

Доповніть твердження, дописавши два слова й формулу:

Невипадкова функція двох аргументів K(t, t'), яка при кожній парі значень t, t' дорівнює кореляційному моменту відповідних перетинів випадкової функції називається ____ і визначається виразом

____.

1
Завдання № 164 **Виберіть правильну відповідь, обвівши кружком цифру** Якщо аргументи кореляційної функції K_x збігаються, тобто t = t', то вона обертається 1) у дисперсію; 2) у математичне сподівання.

1
Завдання № 165 **Виберіть правильну відповідь, обвівши кружком цифру** Як основні характеристики випадкової функції досить розглядати її 1) дисперсію й математичне сподівання; 2) математичне сподівання і кореляційну функцію; 3) дисперсію і кореляційну функцію.

Загальна кількість балів становить 467.

Шкала перекладу балів в оцінки.

Кількість балів	до 115	116-230	231-345	понад 346
Оцінка в п'ятибальній системі	2	3	4	5

Варіанти завдань для контрольної роботи

У процесі вивчення курсу «Теорія імовірностей і математична статистика» студент повинен виконати контрольну роботу, що включає 7 завдань, які охоплюють різні теми курсу. Номер варіанта контрольного завдання вибирають за двома останніми цифрами номера залікової книжки. Всього варіантів - 15. Якщо дві останні цифри залікової книжки перевищують число 15, то номер варіанта визначають шляхом вирахування числа, кратного 15, від 15 до 90. Наприклад, номеру залікової книжки, що закінчується цифрами 86, відповідає варіант 11. Контрольна робота повинна бути виконана в строки, передбачені навчальним графіком. У процесі виконання завдання необхідно наводити відповідні пояснення. Наприкінці роботи слід навести список літератури, яку студент використовував при виконанні контрольної роботи.

На титульному аркуші необхідно написати назву дисципліни, варіант завдання, прізвище, ім'я та по батькові, вказати курс, спеціальність і факультет.

Завдання 1. Для вирішення задач рекомендується використовувати теореми додавання і множення, формулу повної імовірності, теорему гіпотез.

1. Вироби виготовляють два підприємства. У магазин надходить 60% виробів з першого підприємства і 40% - із другого. Перше підприємство випускає 90% виробів без браку і 10% бракованих, а друге - 80% виробів без браку і 20% - бракованих. Знайти імовірність того, що навмання куплений виріб виявиться а) без браку; б) бракованим.
2. На склад надходить продукція трьох фабрик. Продукція першої фабрики становить 25%, другої-40%, третьої - 35%. Відомо також, що імовірність браку для першої фабрики - 4%, для другої - 1% і для третьої -3%. Знайти імовірність того, що обраний навмання виріб: а) стандартний; б) бракований і вироблений на першій фабриці.
3. Булки, які випікає хлібозавод, мають такий розподіл за вагою: менше 90 г - 5%, більше 110 - 10%, інші 85% булок мають нормальну масу (90....110 г). З великої партії беруть навмання дві булки. Знайти імовірність того, що а) обидві булки мають нормальну масу; б) одна булка має масу менше норми, а інша - більше.
4. Працюють три пристрої. Імовірність того, що протягом одного дня перший пристрій відмовить - 0,3, другий - 0,6, третій - 0,1. Знайти імовірність того, що протягом одного дня відмовлять: а) всі пристрої; б) будь-який один; в) принаймні, один пристрій.
5. На столі в певному порядку лежать 32 екзаменаційних білети. Знайти імовірність того, що номер взятого навмання білета буде числом, кратним 5 або 2.
6. Чотири студенти здають іспит. Імовірність того, що перший студент здасть іспит, дорівнює 0,95, другий - 0,9, третій - 0,85, четвертий - 0,8. Знайти імовірність того, що а) хоча б два студенти здадуть іспит; б)

всі чотири студенти здадуть іспит.

7. Достатня умова здачі колоквіуму - відповідь на одне з двох питань, що пропонує викладач студенту. Студент не знає відповідей на десять питань із сорока, які можуть бути запропоновані. Знайти імовірність здачі колоквіуму.
8. Є дві партії виробів по 12 і 10 штук, причому в кожній партії один виріб бракований. Виріб, узятий навмання з першої партії, перекладений в другу, після чого вибирають навмання виріб із другої партії. Визначити імовірність добування бракованого виробу з другої партії.
9. Для контролю продукції з трьох партій деталей взята для випробування одна деталь. Яка імовірність виявлення браку, якщо в одній партії 2/3 деталей браковані, а в двох інших - всі доброякісні?
10. В ящику перебувають 15 тенісних м'ячів, з яких 9 нових. Для першої гри навмання беруть три м'ячі, які після гри повертають в ящик. Для другої гри також навмання беруть три м'ячі. Знайти імовірність того, що всі м'ячі, взяті для другої гри, нові.
11. У тирі є п'ять рушниць, імовірності влучення з яких рівні відповідно 0,5; 0,6; 0,7; 0,8 і 0,9. Визначити імовірність влучення при одному пострілі, якщо стрілець бере одну з рушниць навмання.
12. Є десять однакових урн, з яких в дев'яти перебувають по дві чорних і по дві білих кулі, а в одній - п'ять білих і одна чорна куля. З урни, взятої навмання, витягнута біла куля. Яка імовірність, що куля витягнута з урни, що містить п'ять білих куль?
13. Два стрільця, для яких імовірності влучення в мішень рівні відповідно 0,7 і 0,8, роблять по одному пострілі. Визначити імовірність хоча б одного влучення в мішень.
14. Імовірність настання події в кожному досліді однакова і дорівнює 0,2. Досліди проводять один за іншим до настання події. Визначити імовірність того, що прийдеться робити четвертий дослід.
15. В урні 10 червоних і 6 чорних куль. З урни виймаються одна за іншою три кулі. Знайти імовірність того, що серед них буде не більше однієї червоної.

Завдання 2.

Закон розподілу випадкової величини X заданий таблицею (перший рядок - можливі значення X, другий - відповідні їм значення імовірностей). Знайти: а) математичне сподівання; б) дисперсію; в) середнє квадратичне відхилення випадкової величини X.

Номер варіанта	Вихідні дані					
1	x_i	10	12	20	25	30
	p_i	0,1	0,2	0,1	0,2	0,4
2	x_i	8	12	18	24	30
	p_i	0,3	0,1	0,3	0,2	0,1
3	x_i	30	40	50	60	70
	p_i	0,5	0,1	0,2	0,1	0,1
4	x_i	21	25	32	40	50
	p_i	0,1	0,2	0,3	0,2	0,2
5	x_i	10	12	16	18	20
	p_i	0,2	0,2	0,4	0,1	0,1
6	x_i	11	15	20	25	30
	p_i	0,4	0,1	0,3	0,1	0,1
7	x_i	12	16	21	26	30
	p_i	0,2	0,1	0,4	0,2	0,1
8	x_i	13	17	20	27	30
	p_i	0,1	0,2	0,4	0,2	0,1
9	x_i	14	18	23	28	30
	p_i	0,1	0,4	0,3	0,1	0,1
10	x_i	15	19	24	29	30
	p_i	0,1	0,2	0,2	0,1	0,4
11	x_i	13	17	20	27	30
	p_i	0,1	0,4	0,3	0,1	0,1
12	x_i	14	18	23	28	30
	p_i	0,2	0,1	0,4	0,2	0,1
13	x_i	13	17	20	27	30
	p_i	0,4	0,1	0,3	0,1	0,1
14	x_i	10	12	16	18	20
	p_i	0,1	0,2	0,3	0,2	0,2
15	x_i	8	12	18	24	30
	p_i	0,1	0,2	0,1	0,2	0,4

Завдання 3.

Випадкова величина X задана інтегральною функцією (функцією розподілу F(x)). Знайти: а) диференціальну функцію розподілу (щільність імовірностей); б)математичне сподівання і дисперсію X; в) побудувати графіки інтегральної і диференціальної функцій.

1. $F(x)=\begin{cases}0, x\le 2,\\ \frac{1}{9}(x-2)^2, 2\le x\le 5,\\ 1, x>5.\end{cases}$ **2.** $F(x)=\begin{cases}0, x\le 1,\\ \frac{1}{4}(x-1)^2, 1\le x\le 3,\\ 1, x>3.\end{cases}$ **3.** $F(x)=\begin{cases}0, & x\le 5,\\ \frac{1}{6}(x-1)^2, & 5\le x\le 6,\\ 1, & x>6.\end{cases}$

4. $F(x)=\begin{cases}0, & x\le 3,\\ \frac{1}{4}(x-3)^2, & 3\le x\le 5,\\ 1, & x>5.\end{cases}$ **5.** $F(x)=\begin{cases}0, & x\le 1,\\ \frac{1}{9}(x-1)^2, & 1\le x\le 4,\\ 1, & x>4.\end{cases}$ **6.** $F(x)=\begin{cases}0, & x\le 3,\\ \frac{1}{9}(x-1)^2, & 3\le x\le 6,\\ 1, & x>6.\end{cases}$

7. $F(x)=\begin{cases}0, & x\le 0,\\ \frac{1}{25}x^2, & 0<x\le 5,\\ 1, & x>5.\end{cases}$ **8.** $F(x)=\begin{cases}0, & x\le 2,\\ \frac{1}{6}(x-2)^2, & 2\le x\le 4,\\ 1, & x>4.\end{cases}$ **9.** $F(x)=\begin{cases}0, & x\le 3,\\ \frac{1}{6}(x-2)^2, & 3\le x\le 6,\\ 1, & x>6.\end{cases}$

10. $F(x)=\begin{cases}0, & x\le 6,\\ (x-6)^2, & 6\le x\le 7,\\ 1, & x>7.\end{cases}$ **11.** $F(x)=\begin{cases}0, & x\le 8,\\ (x-8)^2, & 8<x\le 9,\\ 1, & x>9.\end{cases}$ **12.** $F(x)=\begin{cases}0, & x\le 5,\\ (x-8)^2, & 5<x\le 8,\\ 1, & x>8.\end{cases}$

13. $F(x)=\begin{cases}0, & x\le 3,\\ \frac{1}{25}(x-3)^2, & 3<x\le 8,\\ 1, & x>8.\end{cases}$ **14.** $F(x)=\begin{cases}0, & x\le 2,\\ \frac{1}{16}(x-2)^2, & 2\le x\le 6,\\ 1, & x>6.\end{cases}$ **15.** $F(x)=\begin{cases}0, & x\le 2,\\ \frac{1}{16}(x-2)^2, & 2\le x\le 6,\\ 1, & x>6.\end{cases}$

Завдання 4.

Задано математичне сподівання m і середнє квадратичне відхилення σ нормально розподіленої випадкової величини X. Знайти імовірність того, що X прийме значення, що належить інтервалу (α, β), і імовірність того, що абсолютна величина відхилення x-m буде менше ε.

Номер варіанта	m	σ	α	β	ε
1	15	2	9	19	3
2	14	4	10	20	4
3	13	4	10	21	2
4	9	3	9	18	5
5	8	4	8	12	8
6	12	5	12	22	10
7	11	4	13	23	6
8	10	8	14	18	2

Номер варіанта	m	σ	α	β	ε
9	7	2	6	10	1
10	6	2	4	12	0,5
11	3	0,3	1,5	2,5	0,25
12	5	0,2	4	5	0,2
13	11	1	10	11	2
14	4	0,5	3	3,5	2
15	12	1,1	11	12	3

Завдання 5.

Випадкова величина X нормально розподілена з відомим середнім квадратичним відхиленням σ, вибірковою середнєю $\tilde{X}$, обсягом вибірки n. Знайти довірчий інтервал для оцінки невідомого математичного сподівання m з довірчою імовірністю β.

Номер варіанта	$\tilde{X}$	σ	n	β
1	0	0,5	35	0,99
2	10	9,2	30	0,95
3	20	8,2	80	0,9
4	75	1,2	160	0,99
5	8	6,5	20	0,95
6	8	0,9	100	0,9
7	95	8,9	130	0,99
8	13	0,5	170	0,95
9	18	4,5	40	0,9
10	19	3,6	96	0,9
11	35	5	100	0,9
12	50	0,5	120	0,95
13	25	1,5	250	0,95
14	75	6	250	0,9
15	100	5	250	0,9

Завдання 6.

За заданим статистичним розподілом вибірки знайти:

а) вибіркову середню $\tilde{X}$;

б) вибіркову дисперсію $\sigma^2_{виб}$;

в) вибіркове середнє квадратичне відхилення $\sigma_{виб}$.

Номер варіанта	Вихідні дані					
1	x_i	10	12	20	25	30
	n_i	5	18	11	1	9
2	x_i	8	12	18	24	30
	n_i	2	20	7	1	5
3	x_i	30	40	50	60	70
	n_i	1	11	3	1	17
4	x_i	21	25	32	40	50
	n_i	8	8	13	2	18
5	x_i	10	12	16	18	20
	n_i	18	3	11	14	19
6	x_i	11	15	20	25	30
	n_i	19	5	18	8	20
7	x_i	12	16	21	26	30
	n_i	17	20	1	8	2
8	x_i	13	17	20	27	30
	n_i	7	9	1	12	1
9	x_i	14	18	23	28	30
	n_i	15	18	6	13	13
10	x_i	15	19	24	29	30
	n_i	13	17	6	4	9
11	x_i	35	44	64	69	78
	n_i	10	12	9	14	7
12	x_i	8	9	12	13	15
	n_i	15	18	6	13	13
13	x_i	156	165	185	190	200
	n_i	2	4	7	3	1
14	x_i	218	260	270	290	318
	n_i	8	8	13	2	6
15	x_i	51	59	65	78	88
	n_i	7	9	12	14	7

Завдання 7.

Знайти рівняння лінійної регресії у на х, $y_x - \overline{y} = r_b \frac{\delta_y}{\delta_x}(x - \overline{x})$. Дані спостережень наведені в кореляційній таблиці.

Варіант 1

X / Y	4	9	14	19	24	29	
10	2	3					5
20		7	3				10
30			2	50	2		54
40			1	10	6		17
50				4	7	3	14
n_x	2	10	6	64	15	3	100

Варіант 2

X / Y	10	15	20	25	30	35	n_y
30	2	6					8
40		4	4				8
50			7	35	8		50
60			2	10	8		20
70				5	6	3	14
n_x	2	10	13	50	22	3	100

Варіант 3

X / Y	15	20	25	30	35	40	n_y
5	4	2					6
10		6	4				10
15			6	45	2		53
20			2	8	6		16
25				4	7	4	15
n_x	4	8	12	57	15	4	100

Варіант 4

X / Y	5	10	15	20	25	30	n_y
20	1	5					6
30		5	3				8
40			9	40	2		51
50			4	11	6		21
60				4	7	3	14
n_x	1	10	12	55	15	3	100

Варіант 5

X / Y	10	15	20	25	30	35	n_y
6	4	2					6
12		6	2				8
18			5	40	5		50
24			2	8	7		17
30				4	7	8	19
n_x	4	8	9	52	19	8	100

Варіант 6

X / Y	5	10	15	20	25	30	n_y
8	2	4					6
12		3	7				10
16			5	30	10		45
20			7	10	8		25
24				5	6	3	14
n_x	2	7	19	45	24	3	100

Варіант 7

X / Y	2	7	12	17	22	27	n_y
10	2	4					6
20		6	2				8
30			3	50	2		55
40			1	10	6		17
50				4	7	3	14
n_x	2	10	6	64	15	3	100

Варіант 8

X / Y	11	16	21	26	31	36	n_y
25	2	4					6
35		6	3				9
45			6	45	4		55
55			2	8	6		16
65				4	7	3	14
n_x	2	10	11	57	17	3	100

Варіант 9

X Y	4	9	14	19	24	29	n_y
8	3	3					6
18		5	4				9
28			40	2	8		50
38			5	10	6		21
48				4	7	3	14
n_x	3	8	49	16	21	3	100

Варіант 10

X Y	5	10	15	20	25	30	n_y
11	4	2					6
21		5	3				8
31			5	45	5		55
41			2	8	7		17
51				4	7	3	14
n_x	4	7	10	57	19	3	100

Варіант 11

X Y	15	20	25	30	35	40	n_y
5	4	3					7
10		5	4				9
15			6	43	2		51
20			2	10	6		18
25				4	7	4	15
n_x	4	8	12	57	15	4	100

Варіант 12

X Y	5	10	15	20	25	30	n_y
8	2	4					6
12		3	8				11
16			5	30	10		45
20			6	10	8		24
24				5	6	3	14
n_x	2	7	19	45	24	3	100

Варіант 13

X Y	10	15	20	25	30	35	n_y
6	4	2					6
12		6	2				8
18			5	40	5		50
24			2	8	7		17
30				4	7	8	19
n_x	4	8	9	52	19	8	100

Варіант 14

X Y	11	16	21	26	31	36	n_y
25	2	4					6
35		6	3				9
45			6	45	4		55
55			2	8	6		16
65				4	7	3	14
n_x	2	10	11	57	17	3	100

Варіант 15

X Y	2	7	12	17	22	27	n_y
10	2	4					6
20		6	2				8
30			3	50	2		55
40			1	10	6		17
50				4	7	3	14
n_x	2	10	6	64	15	3	100

Запитання до заліку

1) Основні поняття і визначення теорії імовірностей.
2) Методи визначення імовірності випадкової події.
3) Як підрахувати імовірність події класичним методом?
4) Закон великих чисел.
5) У чому відмінність між імовірністю і частотою появи події?
6) Імовірність суми випадкових подій.
7) Слідства з теореми додавання імовірностей.
8) Імовірність добутку випадкових подій.
9) Залежні події. Умовна імовірність випадкової події.
10) Формула повної імовірності.
11) Повторні незалежні досліди. Формула Бернуллі.
12) Дайте визначення найімовірнішого числа появ події. Як обчислити найімовірніше число появ події?
13) Чим розрізняються задачі, в яких потрібне застосування локальної і інтегральної граничних теорем Лапласа?
14) Випадкова величина і її закони розподілу.
15) Властивості функції розподілу випадкової величини.
16) Властивості щільності розподілу випадкової величини.
17) Числові характеристики випадкової величини.
18) Моменти випадкової величини.
19) Що характеризують математичне сподівання і дисперсія? Чим зручне застосування замість дисперсії середнього квадратичного відхилення?
20) Біноміальний закон розподілу. Числові характеристики.
21) Закон розподілу Пуассона. Числові характеристики.
22) Нормальний розподіл випадкової величини. Визначення імовірності влучення випадкової величини на задану ділянку значень.
23) Якими параметрами визначається нормальний закон розподілу випадкової величини?
24) Як змінюється графік нормального закону зі зміною середнього квадратичного відхилення випадкової величини?
25) Центральна гранична теорема.
26) Закон рівномірної щільності. Числові характеристики.
27) Експонентний розподіл випадкової величини. Числові характеристики.
28) Що являє собою багатомірна випадкова величина?
29) Закони розподілу системи випадкових величин.
30) Числові характеристики системи двох випадкових величин.
31) Кореляційний момент і коефіцієнт кореляції.
32) Що характеризує кореляційний момент системи двох випадкових величин?
33) Для чого використовують коефіцієнт кореляції?

34) Предмет і завдання математичної статистики. Основні поняття і визначення.

35) Властивості оцінок числових характеристик.

36) У чому відмінність між середнім арифметичним і математичним сподіванням випадкової величини? Що таке оцінка параметра розподілу?

37) Поясніть властивості спроможності і незміщеності оцінок.

38) Точкові й інтервальні оцінки числових характеристик.

39) Чим відрізняються точкова і інтервальна оцінки параметрів розподілу?

40) Визначення довірчого інтервалу і довірчої імовірності для оцінки математичного сподівання.

41) Генеральна і вибіркова сукупності. Числові характеристики варіаційного ряду.

42) В чому полягає розходження між генеральною сукупністю і вибіркою?

43) Визначення закону розподілу за статистичним даними.

44) Вирівнювання статистичних рядів.

45) Поняття статистичної гіпотези. Нульова і альтернативна гіпотези.

46) Область прийняття гіпотези.

47) Що таке статистична гіпотеза?

48) Які критерії застосовують для перевірки статистичних гіпотез?

49) Критерії згоди.

50) Згладжування експериментальних залежностей за методом найменших квадратів.

51) Поясніть, що таке функція регресії.

52) Охарактеризуйте задачі, розв'язувані за допомогою регресійного аналізу.

53) Для чого застосовують метод найменших квадратів?

54) Що таке рівняння регресії? Як перевіряють його адекватність статистичним даним?

55) Перевірка адекватності рівняння регресії. Оцінка значущості коефіцієнтів рівняння регресії.

56) Поясніть, що таке кореляційна залежність? Як вона пов'язана з поняттям регресії?

57) Охарактеризуйте задачі, розв'язувані за допомогою кореляційного аналізу.

58) Що характеризує коефіцієнт кореляції? Як він пов'язаний з коефіцієнтом регресії?

59) Що таке кореляційна таблиця?

60) Які параметри визначають за допомогою кореляційної таблиці?

61) Які значення може приймати коефіцієнт кореляції?

62) Які завдання вирішують за допомогою однофакторного дисперсійного аналізу? Чому даний вид аналізу випадкових даних одержав назву дисперсійного?

63) Дайте визначення випадкового процесу. Що таке реалізація випадкової функції.

64) Які властивості імовірнісних характеристик стаціонарного випадкового процесу?

65) Дайте визначення кореляційної функції. Якими властивостями вона володіє?

66) Поясніть властивість ергодичності стаціонарного випадкового процесу.

67) Перелічіть основні компоненти системи масового обслуговування, що визначають її математичний опис.

68) Класифікація систем масового обслуговування.

69) Дайте визначення стаціонарного потоку подій і регулярного потоку подій. Поняття марківського процесу.

70) Якими параметрами визначається потік заявок? Потік обслуговувань?

71) Які змінні є невідомими в рівняннях Колмогорова? Що таке фінальної імовірності? Поясните їх зміст.

72) З яких елементів складається граф станів? Поясните їх призначення.

73) Як, маючи граф станів, визначити імовірності станів системи $p_i(t)$?

74) Які показники дозволяють визначити ефективність системи масового обслуговування?

75) В яких випадках використовують метод статистичних випробувань? Поясніть суть методу.

СПИСОК ЛІТЕРАТУРИ

Основна література

1. Гмурман В. Э. Теория вероятностей и математическая статистика. -М.: Высш. школа, 1977. - 498 с.
2. Гмурман В.Э. Руководство к решению задач по теории вероятностей и математической статистике. - М.: Высш. школа, 1975. - 330с.

Додаткова література

3. Карасев А.И., Аксютина З.И., Савельева Т.И. Курс высшей математики для экономических вузов. Ч.2. - М.:Высш. школа, 1982. - 320с.
4. Справочник по математике для экономистов (Под редакцией В.И.Ермакова.) - М.:Высш.школа, 1987. - 306с.
5. Вентцель Е. С. Теория вероятностей. - М.: Высшая школа, 1999.
6. Горбань С.Ф., Снижко Н.В. Теория вероятностей и математическая статистика. - К.: МАУП, 1999.
7. http://teorver.ru
8. http://www.artspb.com
9. http://www.matburo.ru
10. http://stud-project.ru
11. http://www.statsoft.ru
12. http://www.alife.narod.ru
13. http://neuro.net.ua
14. Теория статистики с основами теории вероятностей: Учебное пособие для вузов/ Под ред. И.И. Елисеевой.- М.: ЮНИТИ-ДАНА, 2001.- 446 с.

ПРЕДМЕТНИЙ ПОКАЖЧИК

Р

С

У

Ф

Ц

Ч

Щ

ДОДАТКИ

Додаток 1

Таблиця значень функції $\varphi(x) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$

x	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3653	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	0,2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0,0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0,0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0008	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Додаток 2.

Таблиця значень функції $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{t^2}{2}}dt$

X	Φ(x)	X	Φ(x)	X	Φ(x)	X	Φ(x)
0,00	0,0000	0,36	0,1406	0,72	0,2642	1,08	0,3599
0,01	0,0040	0,37	0,1443	0,73	0,2673	1,09	0,3621
0,02	0,0080	0,38	0,1480	0,74	0,2703	1,10	0,3643
0,03	0,0120	0,39	0,1517	0,75	0,2734	1,11	0,3665
0,04	0,0160	0,40	0,1554	0,76	0,2764	1,12	0,3686
0,05	0,0199	0,41	0,1591	0,77	0,2794	1,13	0,3708
0,06	0,0239	0,42	0,1628	0,78	0,2823	1,14	0,3729
0,07	0,0279	0,43	0,1664	0,79	0,2852	1,15	0,3749
0,08	0,0319	0,44	0,1700	0,80	0,2818	1,16	0,3770
0,09	0,0359	0,45	0,1736	0,81	0,2910	1,17	0,3790
0,10	0,0398	0.46	0,1772	0,82	0,2939	1,18	0,3810
0,11	0,0438	0,47	0,1808	0,83	0,2967	1,19	0,3830
0,12	0,0478	0,48	0,1844	0,84	0,2995	1,20	0,3849
0,13	0,0517	0,49	0,1879	0,85	0,3023	1.21	0,3869
0,14	0,0557	0,50	0,1915	0,86	0,3051	1,22	0,3883
0,15	0,0596	0,51	0,1950	0,87	0,3078	1,23	0,3907
0,16	0,0636	0,52	0,1985	0,88	0,3106	1,24	0,3925
0,17	0,0675	0,53	0,2019	0,89	0,3133	1,25	0,3944
0,18	0,0714	0,54	0,2054	0,90	0,3159	1,26	0,3926
0,19	0,0753	0,55	0,2088	0,91	0,3186	1,27	0,3980
0,20	0,0793	0,56	0,2123	0,92	0,3212	1,28	0,3997
0,21	0,0832	0,57	0,2157	0,93	0,3238	1,29	0,4015
0,22	0,0871	0,58	0,2190	0,94	0,3264	1,30	0,4032
0,23	0,0910	0,59	0,2224	0,95	0,3289	1,31	0,4049
0,24	0,0948	0,60	0,2257	0,96	0,3315	1,32	0,4066
0,25	0,0987	0,61	0,2291	0,97	0,3340	1,33	0,4082
0,26	0,1026	0,62	0,2324	0,98	0,3365	1,34	0,4099
0.27	0,1064	0,63	0,2357	0,99	0,3389	1,35	0,4115
0,28	0,1103	0,64	0,2389	1,00	0,3413	1,36	0,4131
0,29	0,1141	0,65	0,2422	1,01	0,3438	1,37	0,4147
0,30	0,1179	0,66	0,2454	1,02	0,3461	1,38	0,4162
0,31	0,1217	0,67	0,2486	1,03	0,3485	1,39	0,4177
0,32	0,1255	0,68	0,2517	1,04	0,3508	1,40	0,4192
0,33	0,1293	0,69	0,2549	1,05	0,3531	1,41	0,4207
0,34	0,1331	0,70	0,2580	1,06	0,3554	1,42	0,4222
0,35	0,1368	0,71	0,2611	1,07	0,3577	1,43	0,4236

Продовження додатка 2

x	*Φ(x)*	*X*	*Φ(x)*	*X*	*Φ(x)*	*X*	*Φ(x)*
1,44	0,4251	1,73	0,4582	2,04	0,4793	2,62	0,4956
1,45	0,4265	1,74	0,4591	2,06	0,4803	2,64	0,4959
1,46	0,4279	1,75	0,4599	2,08	0,4812	2,66	0,4961
1,47	0,4292	1,76	0,4608	2,10	0,4821	2,68	0,4963
1,48	0,4306	1,77	0,4616	2,12	0,4830	2,70	0,4965
1,49	0,4319	1,78	0,4625	2,14	0,4838	2,72	0,4967
1,50	0,4332	1,79	0,4633	2,16	0,4846	2,74	0,4969
1,51	0,4345	1,80	0,4641	2,18	0,4854	2,76	0,4971
1,52	0,4357	1,81	0,4649	2,20	0,4861	2,78	0,4973
1,53	0,4370	1,82	0,4656	2,22	0,4868	2,80	0,4974
1,54	0,4382	1,83	0,4664	2,24	0,4875	2,82	0,4976
1,55	0,4394	1,84	0,4671	2,26	0,4881	2,84	0,4977
1,56	0,4406	1,85	0,4678	2,28	0,4887	2,86	0,4979
1,57	0,4418	1,86	0,4686	2,30	0,4893	2,88	0,4980
1,58	0,4429	1,87	0,4693	2,32	0,4898	2,90	0,4981
1,59	0,4441	1,88	0,4699	2,34	0,4904	2,92	0,4982
1,60	0,4452	1,89	0,4706	2,36	0,4909	2,94	0,4984
1,61	0,4463	1,90	0,4713	2,38	0,4913	2,96	0,4985
1,62	0,4474	1,91	0,4719	2,40	0,4918	2,98	0,4986
1,63	0,4484	1,92	0,4726	2,42	0,4922	3,00	0,49865
1,64	0,4495	1,93	0,4732	2,44	0,4927	3,20	0,49931
1,65	0,4505	1,94	0,4738	2,46	0,4931	3,40	0,49966
1,66	0,4515	1,95	0,4744	2,48	0,4934	3,60	0,499841
1,67	0,4525	1,96	0,4750	2,50	0,4938	3,80	0,499928
1,68	0,4535	1,97	0,4756	2,52	0,4941	4,00	0,499968
1,69	0,4545	1,98	0,4761	2,54	0,4945	4,50	0,499997
1,70	0,4554	1,99	0,4767	2,56	0,4948	5,00	0,499997
1,71	0,4564	2,00	0,4772	2,58	0,4951		
1,72	0,4573	2,02	0,4783	2,60	0,4953		

Додаток 3

Таблиця значень імовірності $P\{\chi^2 > \chi^2_P\}$

χ^2_P	*k*							
	1	2	3	4	5	6	7	8
1	0,3173	0,6065	0,8013	0,9098	0,9626	0,9856	0,9948	0,9982
2	1574	3679	5724	7358	8491	9197	9598	9810
3	0833	2231	3916	5578	7000	8088	8850	9344
4	0455	1353	2615	4060	5494	6767	7798	8571
5	0254	0821	1718	2873	4159	5438	6600	7576
6	0143	0498	1116	1991	3062	4232	5398	6472
7	0081	0302	0719	1359	2206	3208	4289	5366
8	0047	0183	0460	0916	1562	2381	3326	4335
9	0027	0111	0293	0611	1091	1736	2527	3423
10	0016	0067	0186	0404	0752	1247	1886	2650
11	0009	0041	0117	0266	0514	0884	1386	2017
12	0005	0025	0074	0174	0348	0620	1006	1512
13	0003	0015	0046	0113	0234	0430	0721	1119
14	0002	0009	0029	0073	0146	0296	0512	0818
15	0001	0006	0018	0047	0104	0203	0360	0591
16	0001	0003	0011	0030	0068	0138	0251	0424
17	0000	0002	0007	0019	0045	0093	0174	0301
18		0001	0004	0012	0029	0062	0120	0212
19		0001	0003	0008	0019	0042	0082	0149
20		0000	0002	0005	0013	0028	0056	0103
21		0000	0001	0003	0008	0018	0038	0071
22		0000	0001	0002	0005	0012	0025	0049
23		0000	0000	0001	0003	0008	0017	0034
24		0000	0000	0001	0002	0005	0011	0023
25		0000	0000	0001	0001	0003	0008	0016
26		0000	0000	0000	0001	0002	0005	0010
27		0000	0000	0000	0001	0001	0003	0007
28		0000	0000	0000	0000	0001	0002	0005
29		0000	0000	0000	0000	0001	0001	0003
30		0000	0000	0000	0000	0000	0001	0002

Додаток 4

Критичні точки розподілу χ^2

Число ступенів свободи k	Рівень значущості α					
	0,01	0,025	0,05	0,95	0,975	0,99
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	50,9	47,0	43,8	18,5	16,8	15,0

Додаток5

Критичні точки розподілу Стьюдента

Число ступенів свободи	Рівень значущості α (для двосторонньої критичної області)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,31	12,7	31,82	63,7	318,3	637,0
2	2,92	4,30	6,97	9,92	22,33	31,6
3	2,35	3,18	4,54	5,84	10,22	12,9
4	2,13	2,78	3,75	4,60	7,17	8,61
5	2,01	2,57	3,37	4,03	5,89	6,86
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,89	2,36	3,00	3,50	4,79	5,40
8	1,86	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,78
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,80	2,20	2,72	3,11	4,03	4,44
12	1,78	2,18	2,68	3,05	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,14	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,01
17	1,74	2,11	2,57	2,90	3,65	3,96
18	1,73	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,73	2,09	2,53	2,85	3,55	3,85
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,51	3,79
23	1,71	2,07	2,50	2,81	3,49	3,77
24	1,71	2,06		2,80	3,47	3,74
25	1,71	2,06	2,49	2,79	3,45	3,72
26	1,71	2,06	2,48	2,78	3,44	3,71
27	1,71	2,05	2,47	2,77	3,42	3,69
28	1,70	2,05	2,46	2,76	3,40	3,66
29	1,70	2,05	2,46	2,76	3,40	3,66
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	2,02	2,42	2,70	3,31	3,55
60	1,67	2,00	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,17	3,37
∞	1,64	1,96	2,33	2,58	3,09	3,29
	0,05	0,025	0,01	0,005	0,001	0,0005
	Рівень значущості α (для однобічної критичної області)					

Додаток 6

Значення критерію Фішера при α= 0,01

k_1 / k_2	1	2	3	4	5	6	7	8	9	10	11	12
1	4052	4999	5403	5625	5764	5889	5928	5981	6022	6056	6082	6106
2	98,49	99,01	99,17	99,25	99,33	99,30	99,34	99,36	99,36	99,40	99,41	99,42
3	34,12	30,81	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,54	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6,62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71
11	9,85	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,40
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,61	3,55
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,45

Навчальне видання

АЧКАСОВ Анатолій Єгорович,
ПЛАКІДА Віктор Тарасович,
ВОРОНКОВ Олексій Олександрович,
ВОРОНКОВА Тетяна Борисівна

ТЕОРІЯ ІМОВІРНОСТЕЙ І МАТЕМАТИЧНА СТАТИСТИКА

Навчальний посібник

Редактор М.З. Аляб'єв
Комп'ютерна верстка Т.Б.Воронкова
Дизайн обкладинки Т.Є.Клочко

План 2008, поз. 116 Н

Підп. до друку 24.06.2008	Формат 60 x 84 1/16.	Папір офісний.
Друк на ризографі.	Умовн.-друк.арк. 11,6.	Обл.-вид.арк. 12,0
Зам.№	Тираж 500 прим.	

61002, Харків, ХНАМГ, вул. Революції,12

Сектор оперативної поліграфії ІОЦ ХНАМГ
61002, Харків, ХНАМГ, вул. Революції, 12